

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 13:11:21 ; Search time 1906 Seconds

(without alignments)
5725.893 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375

Sequence: 1 atgaagatattcttgatttc.....atgaagtgtgaagattaa 375

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vl:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vl:*

30: em.htg.hum:*

31: em.htg.iny:*

32: em.htg.other:*

33: em.htg.mus:*

34: em.htg.pln:*

35: em.htg.rod:*

36: em.htg.mam:*

37: em.htg.vrt:*

38: em.sy:*

39: em.htgo.hum:*

40: em.htgo.mus:*

41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	375	100.0	375	6	AX036541	AX036541 Sequence
2	48.6	13.0	189967	2	AL833802	AL833802 Mus muscu
3	48.6	13.0	206914	10	AL807815	AL807815 Mouse DNA
4	46.4	12.4	40090	3	AC006611	AC006611 Caenorhab
5	46.4	12.4	299670	2	AC006845	AC006845 Caenorhab
6	45.8	12.2	4458	2	AC115614	AC115614 Dictyoste
7	45.8	12.2	49999	6	AX015902	AX015902 Sequence
8	45.8	12.2	49999	6	AX015908	AX015908 Sequence
9	45.8	12.2	119191	9	AF004782	AF004782 Homo sapi
10	45.8	12.2	228283	10	AF312994	AF312994 Mus muscu
11	45.2	12.1	20796	3	AF216973	AF216973 Drosophila
12	45.2	12.1	87767	2	AC014497	AC014497 Drosophila
13	45.2	12.1	132000	2	AC116976	AC116976 Dictyoste
14	45.2	12.1	167062	2	AC007624	AC007624 Drosophila
15	45.2	12.1	181771	3	AC008340	AC008340 Drosophila
16	45.2	12.1	348873	3	AE003790	AE003790 Drosophila
17	44.8	11.9	26671	3	CER11A5	283122 Caenorhabd
18	44.4	11.8	2320	9	BC002875	BC002875 Homo sapi
19	44.4	11.8	3230	9	BC010457	BC010457 Homo sapi
20	44.4	11.8	86945	2	AC002490	AC002490 Homo sapi
21	44.4	11.8	153477	2	AC006278	AC006278 Plasmodiu
22	44.4	11.8	154470	2	AC027820	AC027820 Homo sapi
23	44.4	11.8	195770	2	AC091153	AC091153 Homo sapi
24	44.2	11.8	247175	2	AC129593	AC129593 Mus muscu
25	43.6	11.6	110000	2	PFMAL13P2_1	Continuation (2 of
26	43.6	11.6	152409	2	PFMAL1P1	AL031744 Plasmodiu
27	43.6	11.6	183767	2	AL807765	AL807765 Mus muscu
28	43.6	11.6	234295	2	AL714007	AL714007 Mus muscu
29	43.4	11.6	98734	2	PFMAL1P2	AL031745 Plasmodiu
30	43.4	11.6	143716	2	AL627248	AL627248 Dario rer
31	43.4	11.6	168083	9	AC092376	AC092376 Homo sapi
32	43.2	11.5	977	11	CNS061DX	AL400171 T7 end of
33	43.2	11.5	139665	2	AP000865	AP000865 Homo sapi
34	43	11.5	110000	2	PFMAL4P1_1	Continuation (2 of
35	43	11.5	162261	2	AC119001	AC119001 Rattus no
36	42.8	11.4	142439	9	AC008749	AC008749 Homo sapi
37	42.6	11.4	59762	8	AB023032	AB023032 Arabidops
38	42.6	11.4	173016	2	AC019123	AC019123 Homo sapi
39	42.6	11.4	262428	2	AC096326	AC096326 Rattus no
40	42.4	11.3	2783	3	AF093702	AF093702 Plasmodiu
41	42.4	11.3	49144	2	AC131380	AC131380 Strongylo
42	42.4	11.3	115758	9	AC104634	AC104634 Homo sapi
43	42.4	11.3	198516	2	AC102659	AC102659 Mus muscu
44	42.4	11.3	202521	2	AL773546	AL773546 Mus muscu
45	42.4	11.3	234112	3	PFMAL4P2	AL035475 Plasmodiu

ALIGNMENTS

RESULT 1

AX036541

LOCUS

DEFINITION

AX036541

ACCESSION

VERSION

AX036541.1 GI:11226151

KEYWORDS

SOURCE

ORGANISM

Hirudo medicinalis

Eukaryota; Metazoa; Annelida; Hirudinea; Hirudinidae; Hirudo.

REFERENCE

1 (bases 1 to 375)

Hemberger,J., Scheuble,B., Strittmatter,W., Hofmann,U., Fotev,Z. and Gnessow,D.

TITLE

Protein for blocking platelet adhesion

JOURNAL

Patent: WO 0056885-A 1 28-SEP-2000;
HEMBERGER JURGEN (DE) ; MERCK PATENT GMBH (DE) ; SCHEUBLE BERNHARD
(DE) ; STRITTMATTER WOLFGANG (DE) ; HORMANN UWE (DE) ; FOTEV ZIST
(DE) ; GUSSOW DETLEF (DE)

FEATURES

Location/Qualifiers
1..375

CDS

/organism="Hirudo medicinalis"
/db_xref="taxon:6421"
64..375
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC16481.1"
/db_xref="GI:11226152"
/translation="EEREDCMTFVANKRYDPFKSSDLDECKTCKTECYIV
FEDTVNKECYNVVDGEDELQEKVVDENTENVLIDCEKDKAGNAGTDESDYDVE
D"
BASE COUNT 123 a 62 c 85 g 105 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 375; DB 6; Length 375;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCTATTCTTGATTCCTCTGCTTGCCTCGCAAGCTTGCTGATCTCAACTTCT 60
DB 1 ATGAGCTATTCTTGATTCCTCTGCTTGCCTCGCAAGCTTGCTGATCTCAACTTCT 60
QY 61 TCAGAGAAGCTGAAGATTGTTGACGTTTACGGAACAGAAATATACAGACTTCGAT 120
DB 61 TCAGAGAAGCTGAAGATTGTTGACGTTTACGGAACAGAAATATACAGACTTCGAT 120
QY 121 AAATCTTTAAGAACTCCTCGATCTTGACGAATGCAAAAAAACATGTTTCAAGACGAG 180
DB 121 AAATCTTTAAGAACTCCTCGATCTTGACGAATGCAAAAAAACATGTTTCAAGACGAG 180
QY 181 TACTGCTCATCGTTTGAAGACACGCTGACAAAGAAATTTCTACATGCTGTTGAT 240
DB 181 TACTGCTCATCGTTTGAAGACACGCTGACAAAGAAATTTCTACATGCTGTTGAT 240
QY 241 GGTGAGAGTTAGACCAAGAAATTTGTTGTCACGCAAAAACCTCAGGAAATTTATTTG 300
DB 241 GGTGAGAGTTAGACCAAGAAATTTGTTGTCACGCAAAAACCTCAGGAAATTTATTTG 300
QY 301 ACAGACTCGGAGGTAAGATGCGATATGCGGAGGTACAGGTGACGATGAA 360
DB 301 ACAGACTCGGAGGTAAGATGCGATATGCGGAGGTACAGGTGACGATGAA 360
QY 361 GTTGATGAAGATTAA 375
DB 361 GTTGATGAAGATTAA 375

RESULT 2

AL833802 189967 bp DNA linear HTG 17-JUL-2002

LOCUS

Mus musculus chromosome 4 clone RP23-17A4, ** SEQUENCING IN

DEFINITION

Mus musculus chromosome 4 clone RP23-17A4, ** SEQUENCING IN

ACCESSION

AL833802.3 GI:21911787

VERSION

HTG: HTGS PHASE1.

KEYWORDS

Mus musculus.

SOURCE

Mus musculus.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

Submitted (16-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; clone requests: clonerequests@sanger.ac.uk

AUTHORS

On Jul 19, 2002 this sequence version replaced gi:21738804.

TITLE

----- Genome Center
Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk

----- Project Information
Center project name: BM17A4

----- Summary Statistics

Assembly program: XGAP4; version 4.5
Consensus quality: 186623 bases at least Q40
Consensus quality: 187480 bases at least Q30
Consensus quality: 188086 bases at least Q20
Insert size: 188767; sum-of-contigs
Insert size: 215919; 10.9% error; agarose-fp
Quality coverage: 5.28x in Q20 bases; sum-of-contigs quality
coverage: 4.68x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 8454: contig of 8454 bp in length
* 8455 8554: gap of 100 bp
* 8555 19307: contig of 10753 bp in length
* 19308 19407: gap of 100 bp
* 19408 27097: contig of 7690 bp in length
* 27098 27197: gap of 100 bp
* 27198 40822: contig of 13625 bp in length
* 40823 40922: gap of 100 bp
* 40923 48761: contig of 7839 bp in length
* 48762 48861: gap of 100 bp
* 48862 94932: contig of 46071 bp in length
* 94933 95032: gap of 100 bp
* 95033 109692: contig of 14660 bp in length
* 109693 109792: gap of 100 bp
* 109793 122076: contig of 12284 bp in length
* 122077 122176: gap of 100 bp
* 122177 130769: contig of 8593 bp in length
* 130770 130869: gap of 100 bp
* 130870 146267: contig of 15398 bp in length
* 146268 146367: gap of 100 bp
* 146368 151984: contig of 5617 bp in length
* 151985 152084: gap of 100 bp
* 152085 184112: contig of 32028 bp in length
* 184113 184212: gap of 100 bp
* 184213 189967: contig of 5755 bp in length.

FEATURES

source

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-17A4"
/clone_lib="RPCI-23"
1..8454
/note="assembly_fragment:01243
fragment_chain:1
clone_end:SP6
vector_side:left"
8555..19307
/note="assembly_fragment:00590
fragment_chain:1"
19408..27097
/note="assembly_fragment:00259
fragment_chain:1"
27198..40822
/note="assembly_fragment:01468
fragment_chain:1"
40923..48761
/note="assembly_fragment:00793
fragment_chain:2"

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature      48862..94932
                  /note="assembly-fragment:00456
                  fragment.chain:2"
misc_feature      95033..109692
                  /note="assembly-fragment:00530
                  fragment.chain:2"
misc_feature      109793..122076
                  /note="assembly-fragment:01290
                  fragment.chain:2"
misc_feature      122177..130769
                  /note="assembly-fragment:01504.0"
                  /note="assembly-fragment:01504.1"
misc_feature      130870..146267
                  /note="assembly-fragment:01504.1"
misc_feature      146368..151984
                  /note="assembly-fragment:00822
                  fragment.chain:3"
misc_feature      152085..184112
                  /note="assembly-fragment:00663
                  fragment.chain:3"
misc_feature      184213..189967
                  /note="assembly-fragment:00178
                  fragment.chain:3
                  clone_end:17
                  vector_side:right"

```

BASE COUNT 59392 a 34302 c 33745 g 61328 t 1200 others

Query Match 13.0%; Score 48.6; DB 2; Length 189967;
 Best Local Similarity 48.4%; Pred. No. 0.065;
 Matches 135; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```

QY 97 AACGAAATATACAGACTTCGATTAATCTTTAAGAAAGTCTGATCTGACGAATGC 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88182 AACAAACGCCGCAATATTGAACAGTCTTAACAATAAAGAACTTTGGGGAATC 88241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 157 AAAAAAATGTTTCAAGACGAGTACGCTACATCGTTTGAAGACACGGTCAACAAG 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88242 ACCATCCATGACTTCAAGCTATACGACAGCAATGATGATGAGAAAGAAAGAAAGA 88301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 217 GAATGTACTACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88302 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 88361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 277 GAAACCTTACGGAATAATTTATTTGACAGACTCGAGGGTAAAGATGCAAGTATCGGCA 336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88362 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 88421
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 337 GGTACAGGTGACGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88422 GATTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 88460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 3
 AL807815/c 206914 bp DNA linear ROD 22-AUG-2002
 LOCUS Mouse DNA sequence from clone RP23-207H16 on chromosome 4, complete
 DEFINITION sequence.
 ACCESSION AL807815
 VERSION AL807815.3 GI:22474427
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 206914)
 AUTHORS Leonamortier, D.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Aug 23, 2002 this sequence version replaced gi:21690116.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute

COMMENT

Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 chemistries were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-207H16 is
 from the RP23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.choil.org/bacpac/home.htm>
 VECTOR: PBAC3.6.

FEATURES
 Source Location/Qualifiers
 1..206914
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-207H16"
 /clone_lib="RP23-207H16"

BASE COUNT 64113 a 37887 c 37484 g 67430 t

Query Match 13.0%; Score 48.6; DB 10; Length 206914;
 Best Local Similarity 48.4%; Pred. No. 0.067;
 Matches 135; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```

QY 97 AACGAAATATACAGACTTCGATTAATCTTTAAGAAAGTCTGATCTGACGAATGC 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205163 AACAAACGCCGCAATATTGAACAGTCTTAACAATAAAGAACTTTGGGGAATC 205104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 157 AAAAAAATGTTTCAAGACGAGTACGCTACATCGTTTGAAGACACGGTCAACAAG 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205103 ACCATCCATGACTTCAAGCTATACGACAGCAATGATGATGATGATGATGATGATG 205044
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 217 GAATGTACTACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205043 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 204984
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 277 GAAACCTTACGGAATAATTTATTTGACAGACTCGAGGGTAAAGATGCAAGTATCGGCA 336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204983 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 204924
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 337 GGTACAGGTGACGAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204923 GATTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 204885
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 4
 AC006611 40090 bp DNA linear INV 19-APR-2002
 LOCUS Caenorhabditis elegans cosmid G30F8, complete sequence.
 DEFINITION AC006611
 ACCESSION AC006611
 VERSION AC006611.1 GI:4263327
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 40090)

AUTHORS
 TITLE
 Genome sequence of the nematode *C. elegans*: a platform for
 investigating biology. The *C. elegans* Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 JOURNAL
 MEDLINE
 PUBMED
 9851916
 2 (bases 1 to 40090)
 Latreille, P., Twyman, B. and Wilson, R.
 The sequence of *C. elegans* cosmid C30F8
 Unpublished (2001)
 3 (bases 1 to 40090)
 Waterston, R.H.
 Direct Submission
 Submitted (23-FEB-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 40090)
 Waterston, R.H.
 Direct Submission
 Submitted (01-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 40090)
 Waterston, R.
 Direct Submission
 Submitted (14-AUG-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 40090)
 Waterston, R.
 Direct Submission
 Submitted (18-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 40090)
 Waterston, R.
 Direct Submission
 Submitted (12-FEB-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 40090)
 Waterston, R.
 Direct Submission
 Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: rw@nematode.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
[\[www.wormbase.org/db/seq/sequence?name=C30F8;class=Sequence\]](http://www.wormbase.org/db/seq/sequence?name=C30F8;class=Sequence)

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46H11, 8300 bp overlap; the 3' cosmid is Y110A7A,

FEATURES

source

200 bp overlap. Actual start of this cosmid is at base position 197
 of C30F8; actual end is at 40090 of C30F8.
 NOTES:
 Coding sequences below are the result of integration and manual
 review of the following data: computer analysts using the program
 GeneFinder (P. Green and U. Hillier, personal communication), the
 large scale EST projects of Yuji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The *C.*
elegans ORFome cloning project (<http://wormfdb.dfci.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with *C. briggsae*
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual *C. elegans* Genbank submissions,
 and personal communications with *C. elegans* researchers. tRNAs
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers

1..40090

/organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="I"
 /clone="C30F8"
 complement(19565..21512)

gene

complement(19565..21512)
 /note="for a graphical representation of this gene see:
 [www.wormbase.org/db/seq/sequence?name=C30F8.2;class=Seque
 nce]"

CDS

complement(join(19565..19654,20030..20101,20153..20407,
 20460..20617,20879..21057,21102..21273,21392..21512))
 /note="coded for by the following *C. elegans* cDNAs:
 YK71092.3, YK44942.3, YK70591.1, YK74547.3, YK16956.3,
 YK80940.3, YK80740.4, YK70293.3, YK10599.3, YK10599.5,
 YK16956.5, YK30641.3, YK30641.5, YK43141.3, YK43141.5,
 YK44942.5, YK43966.5, YK59640.5, YK71092.5, YK70293.5,
 YK88840.5, YK74547.5, YK76540.5, YK82140.5, YK80940.5,
 YK73905.5, YK70591.5, YK75441.5, YK80740.5,
 YK75441.3, YK88840.3, YK73905.3, YK82140.3"
 /codon_start=1
 /product="Hypothetical protein C30F8.2"
 /protein_id="AK85455.2"
 /db_xref="GI:16259229"

misc_feature

complement(21534..21535)
 /note="SL1 trans-splice site; see YK80940.5"
 25607..32891
 /gene="C30F8.4"
 /note="for a graphical representation of this gene see:
 [www.wormbase.org/db/seq/sequence?name=C30F8.4;class=Seque
 nce]"

CDS

join(25607..25684,27245..27344,27393..27492,27655..27859,
 28346..28436,28489..28707,28818..28897,29946..30020,
 30070..30315,32038..32325,32760..32891)
 /gene="C30F8.4"
 /note="similar to protein kinases (PKA; pkinase, score;
 183.25); coded for by the following *C. elegans* cDNAs:
 YK73909.5, YK28441.3, YK28441.5, YK46547.5, YK73909.3,
 YK6495.5"
 /codon_start=1
 /product="Hypothetical protein C30F8.4a"
 /protein_id="AK85457.1"
 /db_xref="GI:15150650"

100

	LOCUS	AP004782	119191 bp	DNA	linear	PRI 11-JUL-2002
	DEFINITION	Homo sapiens genomic DNA, chromosome 11 clone:CMB9-103B16, complete sequence.				
	ACCESSION	AP004782				
	VERSION	AP004782..3	GI:21728157			
	KEYWORDS	HTG.				
	SOURCE	Homo sapiens DNA, clone:CMB9-103B16.				
	ORGANISM	Homo sapiens				
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.				
	TITLE	Homo sapiens genomic DNA				
	JOURNAL	Published Only in Database (2002)				
	REFERENCE	2 (bases 1 to 119191)				
	AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.				
	TITLE	Direct Submission				
	JOURNAL	Submitted (25-FEB-2002) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shinto-chou,Tsurumi-Ku, Yokohame, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/, Tel.:81-45-503-9111, Fax:81-45-503-9170)				
	COMMENT	On Jul 10, 2002 this sequence version replaced gi:21328197.				
	FEATURES	Location/Qualifiers				
	SOURCE	1. .119191 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="11" /map="11q" /clone="CMB9-103B16"				
	BASE COUNT	34204 a 26845 c 25557 g 32585 t				
	ORIGIN					
	Query Match	12.2%; Score 45.8; DB 9; Length 119191;				
	Best Local Similarity	53.7%; Pred. No.0.34; Indels 0; Gaps 0;				
	Matches 95; Conservative 0; Mismatches 82;					
QY	199 GAAGACAGCGTCAACAAGAATGTACTACATGTCGTGGTAGGAGAGTTAGACC	258				
Db	102658 GAAGACGAAGAAGCAGCAGCAGAAAGAACGCAGCACGAAGAAGACGACGACGCCA	102717				
QY	259 GAAAAATTGTTCGCGANGAATACTTCAGCGSAAAATTATTGACAGACTCGGAGGCTAA	318				
Db	102718 GAAGACGACGACGACGACGAGGAAGACGACGACGACGACGAGGAAGAGACGACGACGAA	102777				
QY	319 GATCGACGATTAATCGCGCGCAGGTACAGTGTCACGATGTCAGTGAATGATGAAGATTAA	375				
Db	102778 GAAGAAGAAGACGAGCAAGAAGAAGAACAAAGAAGAGAAGAAGAAGAAAGAA	102834				
RESULT 10						
AF312994						
LOCUS	228283 bp	DNA	linear	ROD 17-JAN-2002		
DEFINITION	Mus musculus chromosome 1 clone MMU, complete sequence.					
ACCESSION	AF312994					
VERSION	AF312994.1	GI:18182288				
KEYWORDS	HTG.					
SOURCE	Mus musculus					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
AUTHORS	Rump,A., Hess,U., Galgoczy,P., Wirth,T. and Rosenthal,A.					
TITLE	Genomic sequence from mouse chr. 1					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 228283)					
AUTHORS	Rump,A., Hess,U., Galgoczy,P., Wirth,T. and Rosenthal,A.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-OCT-2000) Genome Analysis, Institute of Molecular					

FEATURES									
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany									
location/Qualifiers									
1..228283									
/organism="Mus musculus"									
/db_xref="taxon:10090"									
/chromosome="1"									
/clone="MML"									
BASE COUNT	63688 a	49622 c	49154 g	65819 t					
ORIGIN									
Query Match	12.2%	Score 45.8;	DB 10;	Length 228283;					
Best Local Similarity	49.8%	Pred. No. 0.35;							
Matches 116; Conservative	0;	Mismatches 117;	Indels 0;	Gaps 0;					
QY 143	ATCTTGACGATGCAAAAAACATGTTTCAAGAGGAGTACTGCTACATCGTTTGTGAAG	202							
Db 148716	ATTTTGAAGAAAAAAATCATATATTACAGATTCCAGACTGTTATATGAAACAAG	148775							
QY 203	ACACGGTCACACAGGATGTTACTACATGCTGCTGATGGTGAAGAGTTAGCCACAAA	262							
Db 148776	AAGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	148835							
QY 263	AATTTGTGTCGACGAAAACCTTCACGGAATTTATTTTCACAGCTCGAGGCTAAAGATG	322							
Db 148836	AAGAAG	148895							
QY 323	CAGTAAATGCGCGCAGGTACAGGTGACGATCGATGGAAGTTGATGAAGATTAA	375							
Db 148896	AAGAAG	148948							
RESULT 11									
AF216973		20796 bp	DNA	linear	INV 17-FEB-2000				
LOCUS									
DEFINITION	Drosophila melanogaster 1(2)01289 gene, complete cds, alternatively								
ACCESSION	AF216973								
VERSION	AF216973								
KEYWORDS	AF216973.1 GI:6984064								
SOURCE									
ORGANISM	Drosophila melanogaster.								
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;								
AUTHORS	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;								
TITLE	Ephydriidae; Drosophilidae; Drosophila.								
JOURNAL	1. (bases 1 to 20796)								
REFERENCE	Manning, G., Micklem, D. R. and Krasnow, M. A.								
AUTHORS	Sequence of 1(2)01289 genomic locus								
TITLE	Unpublished								
JOURNAL	2 (bases 1 to 20796)								
REFERENCE	Manning, G., Micklem, D. R. and Krasnow, M. A.								
AUTHORS	Submitted (16-FEB-1999) Department of Biochemistry, Stanford								
TITLE	University, Beckman Center, Stanford, CA 94305, USA								
JOURNAL	Location/Qualifiers								
FEATURES									
source	1..20796								
gene	/organism="Drosophila melanogaster"								
mRNA	/db_xref="taxon:7237"								
	<865..20489								
	/gene="1(2)01289"								
	join(-865..1470,3154..3310,3382..3548,3980..4015,								
	6598..6780,6911..7054,7117..7250,7315..7495,7567..7881,								
	8352..8566,8627..9271,9325..9507,9575..9712,9776..9958,								
	15122..15253,15483..15806,15871..16059,16118..16255,								
	17488..17670,17723..17854,18603..18759,18817..18983,								
	19047..19203,19257..19423,19484..19666,19733..20489)								
	/gene="1(2)01289"								
	/product="1(2)01289 long form"								

	C	10	39.6	10.6	766	22	ABA67193	Human breast cell
	C	11	39.6	10.6	766	22	ABA67193	Human foetal liver
	C	12	39.6	10.6	766	22	ABA4288	Probe #12754 for g
	C	13	39.6	10.6	766	22	AAI12211	Probe #12047 for g
	C	14	39.6	10.6	766	22	AAI17409	Probe #16095 used t
	C	15	39.6	10.6	766	22	AAI07812	Probe #7803 used t
	C	16	39.6	10.6	1944	22	ABAA4123	Human breast cell
	C	17	39.6	10.6	1944	22	ABA54575	Human foetal liver
	C	18	39.6	10.6	1944	22	ABA34359	Probe #2825 for ge
	C	19	39.6	10.6	1944	22	AAI12874	Probe #2807 for ge
	C	20	39.6	10.6	1944	22	AAI14231	Probe #2917 used t
	C	21	39.6	10.6	1944	22	AAI02791	Probe #2782 used t
	C	22	39.6	10.6	3489	21	AAA30290	Kaposi's sarcoma-a
	C	23	39.6	10.6	3489	22	AAF82901	Nucleotide sequenc
	C	24	39.6	10.6	3489	24	ABA93487	Kaposi's sarcoma-a
	C	25	39.6	10.6	32207	20	AAV33805	KSHV LTR DNA (nucl
	C	26	39.6	10.6	137507	19	AAV19941	KSHV Long unique c
	C	27	39.4	10.5	354	23	AAAS90677	DNA encoding novel
	C	28	39.4	10.5	654	23	AAAS69553	DNA encoding novel
	C	29	39.4	10.5	654	23	AAAS71153	DNA encoding novel
	C	30	39.4	10.5	654	23	AAAS75467	DNA encoding novel
	C	31	39.2	10.5	234	23	AAAS75455	DNA encoding novel
	C	32	39	10.4	400	23	AAAS75460	DNA encoding novel
	C	33	39	10.4	963	23	AAAS68580	DNA encoding novel
	C	34	38.6	10.3	1448	21	AAC47605	Arabidopsis thalia
	C	35	38.4	10.2	1072	23	AAAS90738	DNA encoding novel
	C	36	37.8	10.1	2944528	24	ABA03041	Listeria monocytoc
	C	37	37.6	10.0	1622	23	AAAS64269	DNA encoding novel
	C	38	37.4	10.0	2658	21	AAAT0213	Plasmidium falcipa
	C	39	37.4	10.0	3543	22	AAAT8219	Shrimp white spot
	C	40	37.4	10.0	305107	22	AAHS2689	Shrimp white spot
	C	41	37.2	9.9	315	22	ABAA51452	Human breast cell
	C	42	37.2	9.9	315	22	ABAA69479	Human foetal liver
	C	43	37.2	9.9	315	22	ABA36403	Probe #14869 for g
	C	44	37.2	9.9	315	22	AAK17735	Human brain expres
	C	45	37.2	9.9	315	22	AAAK3555	Human bone marrow

RESULT 1

ID AAA28180 standard; CDNA: 375 BP.

XX AAA28180;

DT 29-JAN-2001 (first entry)

DE Platelet binding inhibitor protein Saratin encoding CDNA sequence.

XX Saratin; medicinal leech; thromboembolic disease; intraocular lens;

KW collagen-dependent platelet adhesion inhibition;

KW posterior capsule opacification; ss.

OS Hirudo medicinalis.

XX Key Location/Qualifiers

FT CDS 64..375

FT /*tag= a

FT /product= "Saratin"

FT /note= "Platelet adhesion inhibitor protein"

PX WO200056885-A1.

XX 28-SEP-2000.

PD 10-MAR-2000; 2000WO-EP02117.

PF 18-MAR-1999; 99EP-0105530.

PR 12-MAY-1999; 99EP-0109503.

PA (MERCK) MERCK PATENT GMBH.

ALIGNMENTS

XX Strittmatter W, Gnessow D, Hofmann U, Hemberger J, Fotev Z;
PI Scheuble B;
XX WPI: 2000-611629/58.
DR P-PSDB: AA194746.
XX
XX New Saratin polypeptide and gene isolated from Hirudo medicinalis for
PT blocking platelet adhesion, especially useful for treating or
PT preventing thrombotic diseases, or for manufacturing a medicament for
PT thromboembolic diseases
XX
XX Claim 4; Page 40; 46pp; English.

CC Saratin is a protein isolated from the saliva of the medicinal leech
CC Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
CC platelet adhesion. The invention includes polynucleotide sequences
CC encoding Saratin, an expression vector comprising the DNA sequence, a
CC host cell transformed with the expression vector, antibodies
CC immunospecific for Saratin, and methods for identifying Saratin agonists
CC or antagonists. Saratin is useful for treating thromboembolic processes,
CC and for manufacturing a medicament for treating thromboembolic diseases.
CC It is useful for preventing thrombotic diseases. Saratin is also useful
CC for coating artificial surfaces, since use of Saratin renders them
CC non-adhesive for cells and prevents the activation of cells. It may also
CC be used for coating natural collagen surfaces. Furthermore, Saratin is
CC useful for modifying intraocular lenses in order to lessen the
CC thrombogenicity of the lens material, for contacting the lens surface, or
CC for covalent crosslinking to modify the lens material. The lens material
CC is used for refractive anterior or posterior chamber ocular implants,
CC which may be implanted into the eye. This new type of coating avoids
CC problems contributed by stimulated cell growth. In combination with other
CC medicaments that are for instance conferring cell death, Saratin coating
CC helps to completely overcome posterior capsule opacification. The
CC antibody immunospecific for Saratin, as well as Saratin itself, are
CC useful for measuring samples derived from host cell cultures or from a
CC treated subject. The present sequence represents cDNA encoding the
CC Saratin protein.
XX
XX Sequence 375 BP; 123 A; 62 C; 85 G; 105 T; 0 other;

Query Match 100.0%; Score 375; DB 21; Length 375;
Best Local Similarity 100.0%; Pred. No. 3.3e-99;
Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTATTTCTGATTTCTTCTTCCCTGCGAAGCTTGATCTCAACTCTCT 60
DB 1 ATGAAGTATTTCTGATTTCTTCTTCCCTGCGAAGCTTGATCTCAACTCTCT 60
QY 61 TCAGAAAGACGTGAGATGTTGGACGTTTACGCGAAGAAATATACAGACTTCGAT 120
DB 61 TCAGAAAGACGTGAGATGTTGGACGTTTACGCGAAGAAATATACAGACTTCGAT 120
QY 121 AATCTTTTAAAGAGCTCTGATCTTGACGAATGCAAAAAACATGTTTCAAGCGAG 180
DB 121 AATCTTTTAAAGAGCTCTGATCTTGACGAATGCAAAAAACATGTTTCAAGCGAG 180
QY 181 TACTGCTACATCGTTTAAAGACAGGTCAACAAGATGTTACTACAATCTGTTAT 240
DB 181 TACTGCTACATCGTTTAAAGACAGGTCAACAAGATGTTACTACAATCTGTTAT 240
QY 241 GGTGAAGAGTACACCAAGAAAAATTTGTTGACGAAAACTTCACGAAAAATTTAT 300
DB 241 GGTGAAGAGTACACCAAGAAAAATTTGTTGACGAAAACTTCACGAAAAATTTAT 300
QY 301 ACAGATGCGAGAGGTAAAGATCAGGTAATGCGGACGATACAGGTGACAGATCAAT 360
DB 301 ACAGATGCGAGAGGTAAAGATCAGGTAATGCGGACGATACAGGTGACAGATCAAT 360
QY 361 GTTGATGAGATTTAA 375
DB 361 GTTGATGAGATTTAA 375

RESULT 2
AA223891
ID AA223891 standard; DNA; 49999 BP.
XX
XX
XX AA223891;
XX
XX 25-JAN-2000 (first entry)
XX
XX Murine LOBO genomic DNA fragment 1.
XX

KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.

OS Mus musculus.
PN NC9950284-A2.
XX
XX 07-OCT-1999.
XX
XX 26-MAR-1999; 99MO-EP02055.
XX
XX 27-MAR-1998; 98DE-1013799.

XX (ROSE/) ROSENTHAL A.
XX
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
DR WPI; 1999-601320/51.
XX
XX Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
XX Example 3; Page 69-97; 391pp; German.

XX This invention describes novel nucleic acids (I; designated LOBO (long
XX bones)) encoding proteins influencing bone development in mammals. The
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX development), with exception of the skull and have osteopathic activity.
XX The nucleic acid molecules, proteins and antibodies can be used in
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX animals, especially a transgenic mouse for the study of diseases
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX achondroplasia. This sequence encodes the murine LOBO protein described
XX in the method of the invention.

QY Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;
Query Match 12.2%; Score 45.8; DB 20; Length 49999;
Best Local Similarity 49.8%; Pred. No. 0.011;
Matches 116; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 143 ATCTTGACGAATGCAAAAAACATGTTTCAAGACGAGTCTGCTACATCGTTTGAAG 202
DB 143 ATCTTGACGAATGCAAAAAACATGTTTCAAGACGAGTCTGCTACATCGTTTGAAG 202
QY 203 ACAGGTCACCAAGAGATGTTACTACAATGCGTTGATGGAAGATTAGACCAAGAA 262
DB 203 ACAGGTCACCAAGAGATGTTACTACAATGCGTTGATGGAAGATTAGACCAAGAA 262
QY 263 AATTTGTTGCGAAGAAACTTCACGAAATTTTTCACAGACTGCGAGGTAAGATG 322
DB 263 AATTTGTTGCGAAGAAACTTCACGAAATTTTTCACAGACTGCGAGGTAAGATG 322
QY 323 CAGGTAATGCGGACGTACAGGTGACGATGATGATGATGATGATGATGATGAT 375
DB 323 CAGGTAATGCGGACGTACAGGTGACGATGATGATGATGATGATGATGATGAT 375
QY 7513 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 7565
DB 7513 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 7565

RESULT 3
AA223896

[illegible][illegible]

22-AUG-2002 (first entry)

Human ovarian antigen HNBV053 CDNA, SEQ ID NO:796.

Human: ovarian antigen; ovary; ovarian; breast; cancer; tumour;
ovarian cancer; breast cancer; tumour; reproductive system disorder;
infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
PCOS: ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
inflammatory condition; immune disorder; blood disorder;
cardiovascular disorder; respiratory disorder; neurological disorder;
gastrointestinal disorder; urinary system disorder; drug screening;
gene therapy; chromosome mapping; forensic analysis;
antibody preparation; cytostatic; immunomodulatory; neuroprotective;
antiinflammatory; gynaecological; reproductive; gene; ss.

Homo sapiens.

MO200200677-A1.

03-JAN-2002.

07-JUN-2001; 2001MO-US18569.

07-JUN-2000; 2000US-209467P.

(HUMA-) HUMAN GENOME SCI INC.

Birse CE, Rosen CA;
WPI; 2002-147878/19.
P-PSDB; ABP41839.

Isolated nucleic acid molecules encoding novel ovarian polypeptides,
useful in the prevention, treatment and diagnosis of cancer (e.g.
ovarian cancer). Immune disorders, cardiovascular disorders and
neurological diseases -

Claim 1; SEQ ID NO 796; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-
ABP43228) and to cDNAs encoding them (ABO54131-ABO56305), and also
encompasses polypeptides 90% identical and polynucleotides 95% identical
to the sequences of the invention. The invention additionally relates to
recombinant vectors and host cells comprising human ovarian antigen
polynucleotides, antibodies against human ovarian antigens, and the use
of ovarian antigen polynucleotides and polypeptides in diagnosing,
treating, prognosing or preventing various ovary and/or breast-related
disorders. Such conditions include ovarian cancer and breast cancer, and
metastatic tumours of ovarian or breast origin, reproductive system
disorders (e.g., infertility, disorders of pregnancy, anovulation,
polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
vaginitis), immune disorders (e.g., congenital and acquired
immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus)
blood-related disorders (e.g., anaemia), cardiovascular disorders,
respiratory disorders, neurological disorders, gastrointestinal disorder
and urinary system disorders. Ovarian antigen polypeptides and
polynucleotides may also be used in screening for compounds which
modulate ovarian antigen expression or activity. The polynucleotides may
further be used for gene therapy, chromosome mapping, in the
identification of individuals and in forensic analysis, and the
polypeptides may be used as food additives or to prepare antibodies
useful in disease diagnosis, drug targeting and phenotyping. The present
sequence represents cDNA encoding a human ovarian antigen of the
invention.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at [ftp://wipo.int/pub/published_pcl_sequences](http://wipo.int/pub/published_pcl_sequences).

Sequence 2438 BP; 561 A; 761 C; 686 G; 427 T; 3 other;

11.8%; Score 44.4; DB 24; Length 2438;

[illegible]

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 372 BP; 194 A; 27 C; 134 G; 17 T; 0 other;

Query Match 10.9%; Score 41; DB 23; Length 372;
 Best Local Similarity 48.9%; Pred. No. 0.043;
 Matches 110; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

OY 151 GAATCAAAAAACATGTTCAACGAGGAGTCTGATCATCGTTTGAAGACACGGTC 210
 DB 31 GAATCCCATTAACGGAGATTGAGACTCCTAGGCAACATAGCGAACCCCATCTGAAA 90
 OY 211 AACAAAGAAATGTTACTACAAATGTCGTTGATGTTGAAGAGTTAGACCAAGAAAAATTTGTT 270
 DB 91 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 150
 OY 271 GTGACGAAAACTCAGCGAAATTTATTTGACAGACTGCGAGGTTAAAGATGACGTAAT 330
 DB 151 GAGAGAGAAAGAAAG 210
 OY 331 GCGGACAGGTACAGGTGACGAGTCAAGTCAAGTTGATGAGATTA 375
 DB 211 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 255

RESULT 7
 AAT89346
 ID AAT89346 standard; cDNA; 3211 BP.

AC AAT89346;

DT 11-MAR-1998 (first entry)

XX Human p160 cDNA 160.2.

XX p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 KM p56-lck; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 439..3156

FT /*tag= a
 FT /product= p160
 FT /note= "160.2"

XX W09722255-A1.

XX 26-JUN-1997.

XX 11-DEC-1996; 96WO-US19944.

XX 19-DEC-1995; 95US-0574959.

XX (DAND) DANA FARBER CANCER INST INC.

XX Young I, Shin J, Strominger JL, Vadamudi RK;

XX WPI: 1997-341351/31.

XX P-PSDB; AAW31186.

XX cDNA encoding p62 and p160 and corresponding proteins - used in the
 PT treatment of autoimmune disease and for T and B cell proliferation,
 XX e.g. for treatment of tumours

XX Claim 82; Fig 10; 175pp; English.

CC This cDNA sequence encodes a novel p160 (160.2) which is capable
 CC of activating transcription of a variety of genes upon activation of p62
 CC and is capable of binding to the p62/p56lck complex to modulate lck
 CC function in a manner similar to p62. The genes transcribed in response to
 CC p160 activation likely include those of which are involved in T or B cell
 CC development/differentiation, T or B cell activation or production of T or
 CC B cell specific factors e.g. lymphokines or antibodies. This p160
 CC polypeptide is also a substrate for serine/threonine kinase activity.
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
 CC cycle regulatory proteins stimulating expression of cell cycle dependent
 CC kinase inhibitors and arresting cell cycle progression at specific
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
 CC response it may be used to treat disorders where this is beneficial, e.g.
 CC infections by pathogenic microorganisms. p160 can be used to expand T
 CC cell populations for treating infectious diseases or cancer and p160
 CC inhibitors could reduce B or T cell responses and may be used to treat a
 CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
 CC multiple sclerosis allergic reactions, Crohn's diseases etc.

XX Sequence 3211 BP; 649 A; 990 C; 948 G; 624 T; 0 other;

Query Match 10.7%; Score 40.2; DB 18; Length 3211;
 Best Local Similarity 52.0%; Pred. No. 0.17; 83; Indels 0; Gaps 0;
 Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

OY 199 GAGACACGCTCAACAAGATCTTACTCAATGTCGTTGATGTTGAAGATTAGACCA 258
 DB 2458 GAGACAGAGAAAG 2517
 OY 259 GAAAAATTTGTTGTCAGCAAGAACTTCACGAAAAATTTATTTGACAGACTGCGAGGTTAA 318
 DB 2518 GAATATTTTGAAG 2577
 OY 319 GATGCAAGTAAATCGGACAGGTACAGGTGACGATGATGATGATGATGATGATGAT 371
 DB 2578 GAAAGTGAGTTAG 2630

RESULT 8
 AAT89345
 ID AAT89345 standard; cDNA; 3901 BP.

AC AAT89345;

DT 11-MAR-1998 (first entry)

XX Human p160 cDNA 160.1.

XX p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 KM p56-lck; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 439..3846

FT /*tag= a
 FT /product= p160
 FT /note= "160.1"

XX W09722255-A1.

XX 26-JUN-1997.

XX 11-DEC-1996; 96WO-US19944.

XX 19-DEC-1995; 95US-0574959.

XX (DAND) DANA FARBER CANCER INST INC.

XX Young I, Shin J, Strominger JL, Vadamudi RK;

XX WPI: 1997-341351/31.

XX	New nucleotide sequence useful in the identification or Lactococcus
PT	lactic acid related species -
XX	
PS	Claim 1; SEQ ID 1; 2504bp; French.
XX	
CC	The present invention is related to a Lactococcus lactis nucleotide
CC	sequence (AB990521) and related proteins (AB853300-AB85621). The
CC	nucleic acid sequence is useful in the detection and/or amplification of
CC	nucleic acid sequence, particularly to identify Lactococcus lactis or
CC	related species. The proteins of the invention are useful for the
CC	biosynthesis or biodegradation of a composition of interest. The
CC	invention helps research in lactic bacteria, particularly useful in the
CC	production of yogurt and cheese.
CC	Note: The sequence data for this patent is based on equivalent patent
CC	WO20017734 (published 18-OCT-2001) which is available in electronic
CC	format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX	
SQ	Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;
	Query Match 10.7% Score 40; DB 24; Length 2365589;
	Best Local Similarity 50.5%; Pred. No. 2.4; Mismatches 95; Indels 0; Gaps 0;
	Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY	57 TTCTTCAGAAAGACGTGAAGATGTTGGACGTTTACCGCAAGAAAATTACAGACTT 116
Dy	997986 TTCTTCAGAAAGAATAATGATATTTTCTGCTCTCTTTAAGGATTTAAAGATTACAGATTT 997927
QY	117 CGATTAATCTTTTAAGAAGTGCTCTGATCTTGACGAATGCAGAAAAACATCTTCAAGAC 176
Dy	997926 ACATTTTGCTGCTGATATATTTCAAATGATTTTCCAGAAATTTCCGAAACCATTTTCAAGCA 997867
QY	177 GGAGTACTGCTACATCGCTTTTGAAGACACGCGTCAACAGAGATGTTACTACAATGTCGT 236
Dy	997866 AATGAATTTACTGACAGAGCATTACAGTCACTTACAATCTTGGAAATCACGATATGTTGG 997807
OY	237 TGATGCTGAGCA 248
Dy	997806 TCTTTCGAGCA 997795
RESULT 10	
ID	ABA49279/C
XX	ABA49279 standard; DNA; 766 BP.
XX	
AC	ABA49279;
XX	
DT	01-FEB-2002 (first entry)
XX	
DE	Human breast cell single exon nucleic acid probe #7974.
XX	
KW	Human; microarray; single exon probe; gene expression; breast;
KW	disease; cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157271-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00662.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
P1	Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.
 DR New spatially-addressable set of single exon nucleic acid probes,
 XX useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes
 PT
 PS Claim 4; SEQ ID NO 7974; 327bp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other:
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTGAAGACAGCGTCACACAGATGTTACTACAAATGTCGTGATGTA 246
 Db TCCACACTTGAAGACACCAATGATGATGATGATGATGATGATGATGATGATGATG 628
 QY 247 GAGTTAGCCAAAGAAATTTGTCGACGAAACTTCACGAAATTTTGGACAGAC 306
 Db GATGAGAGAGATGATATATATGATGATGATGATGATGATGATGATGATGATGATG 568
 QY 307 TGCAGAGGTAAGATGACAGTAAATGCGCAGGTACAGGTGACAGTCAATGATGAT 366
 Db GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 Db 507 GA 506
 RESULT 11
 ABA67193/c
 ID ABA67193 standard; DNA; 766 BP.
 AC
 XX ABA67193:
 DT 01-FEB-2002 (first entry)
 DE Human foetal liver single exon nucleic acid probe #15498.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 OS Homo sapiens.
 OS
 PN WO200157277-A2.
 PD
 XX 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US000669.
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 4; SEQ ID NO 15498; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other:
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTGAAGACACGTCACACAGATGTTACTACAAATGTCGTGATGTA 246
 Db TCCACACTTGAAGACACCAATGATGATGATGATGATGATGATGATGATGATGATG 628
 QY 247 GAGTTAGCCAAAGAAATTTGTCGACGAAACTTCACGAAATTTTGGACAGAC 306
 Db GATGAGAGAGATGATATATATGATGATGATGATGATGATGATGATGATGATGATG 568
 QY 307 TGCAGAGGTAAGATGACAGTAAATGCGCAGGTACAGGTGACAGTCAATGATGAT 366
 Db GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 Db 507 GA 506
 RESULT 12
 ABA34288/c
 ID ABA34288 standard; DNA; 766 BP.
 AC
 XX ABA34288:
 DT 23-JAN-2002 (first entry)
 DE Probe #12754 for gene expression analysis in human heart cell sample.
 XX
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 OS Homo sapiens.
 OS
 PN WO200157274-A2.
 PD
 XX 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00066.
 PR 04-FEB-2000; 2000US-0180312.

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 16:39:12 ; Search time 50 Seconds
(without alignments)
2300.076 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagattcttctgatttc.....atgaagtgtgaagatttaa 375

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.2	11.3	7218	1 US-08-232-463-14	Sequence 14, Appl
2	40.2	10.7	3211	2 US-08-574-959A-8	Sequence 8, Appl1
3	40.2	10.7	3211	4 US-09-357-014-8	Sequence 8, Appl1
4	40.2	10.7	3901	2 US-08-574-959A-6	Sequence 6, Appl1
5	40.2	10.7	3901	4 US-09-357-014-6	Sequence 6, Appl1
6	39.6	10.6	3489	2 US-08-728-323A-1	Sequence 1, Appl1
7	39.6	10.6	3489	4 US-09-298-568-1	Sequence 1, Appl1
8	39.6	10.6	32207	2 US-08-770-379-20	Sequence 20, Appl1
9	39.6	10.6	32207	4 US-08-757-669A-20	Sequence 20, Appl1
C 10	39.6	10.6	32207	4 US-09-230-371A-20	Sequence 20, Appl1
11	36	9.6	966	2 US-08-766-738-2	Sequence 2, Appl1
12	36	9.6	966	4 US-09-262-610-2	Sequence 2, Appl1
13	35	9.3	2694	2 US-08-867-941-2	Sequence 2, Appl1
14	35	9.3	2694	4 US-08-074-658-2	Sequence 2, Appl1
15	35	9.3	7650	2 US-08-867-941-1	Sequence 1, Appl1
16	35	9.3	7650	4 US-09-074-658-1	Sequence 1, Appl1
17	35	9.3	8920	2 US-08-446-855A-1	Sequence 1, Appl1
18	35	9.3	8920	4 US-09-150-741-1	Sequence 1, Appl1
19	34.8	9.3	5361	4 US-08-973-462-2	Sequence 2, Appl1
20	34.8	9.3	6152	4 US-08-973-462-1	Sequence 1, Appl1
21	34.6	9.2	2518	4 US-09-433-699-3	Sequence 3, Appl1
22	34.2	9.1	5852	1 US-07-867-106-2	Sequence 2, Appl1
23	34	9.1	759	1 US-08-466-603-4	Sequence 4, Appl1
24	34	9.1	759	1 US-08-314-503A-4	Sequence 4, Appl1
25	34	9.1	759	1 US-08-468-066-4	Sequence 4, Appl1
26	34	9.1	759	2 US-08-466-717-4	Sequence 4, Appl1
27	34	9.1	759	3 US-08-466-743-4	Sequence 4, Appl1

ALIGNMENTS

28	34	9.1	759	5 PCT-US95-12414-4	Sequence 4, Appl1
29	34	9.1	980	1 US-08-466-603-3	Sequence 3, Appl1
30	34	9.1	980	1 US-08-314-503A-3	Sequence 3, Appl1
31	34	9.1	980	1 US-08-468-066-3	Sequence 3, Appl1
32	34	9.1	980	2 US-08-466-717-3	Sequence 3, Appl1
33	34	9.1	980	3 US-08-466-743-3	Sequence 3, Appl1
34	34	9.1	980	5 PCT-US95-12414-3	Sequence 3, Appl1
35	33.8	9.0	4518	4 US-08-961-527-121	Sequence 121, App
36	33.6	9.0	2295	1 US-08-375-300-3	Sequence 3, Appl1
37	33.6	9.0	2295	3 US-09-177-431-3	Sequence 3, Appl1
38	33.6	9.0	2295	5 PCT-US95-16930-3	Sequence 3, Appl1
39	33.6	9.0	4080	1 US-08-375-300-1	Sequence 1, Appl1
40	33.6	9.0	4080	3 US-09-177-431-1	Sequence 1, Appl1
41	33.6	9.0	4080	5 PCT-US95-16930-1	Sequence 1, Appl1
42	33.4	8.9	1137	4 US-09-134-001C-657	Sequence 657, App
43	33.4	8.9	1727	4 US-09-071-035-295	Sequence 295, App
44	33.4	8.9	1839	4 US-09-071-035-293	Sequence 293, App
45	33.2	8.9	4599	1 US-08-431-080-27	Sequence 27, Appl

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14
Query Match 11.3%; Score 42.2; DB 1; Length 7218;

Best Local Similarity 4.6%; Pred. No. 0.0023;
Matches 14; Conservative 170; Mismatches 123; Indels 0; Gaps 0;

QY 65 AGACAGTGAAGATTGTTGACGTTTACGCAACAGAAATATACAGCTTGTGTAAT 124
||||| ||||| :
Db 1456 AAGAGTGAAGATTGTTGACGTTTACGCAACAGAAATATACAGCTTGTGTAAT 1397
QY 125 CTTTGAAGTGTCTGCTGTTGACGTTTACGCAACAGAAATATACAGCTTGTGTAAT 184
:
Db 1396 RRR 1337
QY 185 GTCATACGTTTGAAGACGAGTCAACAGAAATGTTACTACATGCTGTGATGCTG 244
:
Db 1336 RRR 1277
QY 245 AAGAGTGAAGATTGTTGACGTTTACGCAACAGAAATATACAGCTTGTGTAAT 304
:
Db 1276 RRR 1217
QY 305 ACTGCGAGGTAAAGATGAGTAAATGCGGACGATGACGATGAGTGAAGTTG 364
:
Db 1216 RRR 1157
QY 365 ATGAGA 371
:
Db 1156 RRRRRR 1150

RESULT 2

US-08-574-959A-8
; Sequence 8, Application US/08574959A
; Patent No. 5962224

GENERAL INFORMATION:
APPLICANT: Jaekyoon Shin, Insil Jouny, Ratna K. Vadlamudi
and Jack L. Strominger
TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/574,959A
FILING DATE: 19-DEC-95
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 439..3157
US-08-574-959A-8

Query Match 10.7%; Score 40.2; DB 2; Length 3211;

Best Local Similarity 52.0%; Pred. No. 0.0066;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 199 GAACACGCGTCAACAGAAATGTTACTACATCTCTGATGTTGAGAGTTAGACCA 258
||||| ||||| :
Db 2458 GAACAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2517
QY 259 GAAATATTTGTTGCGAGAAACTTCAAGAAATATTTGACAGCTCGAGGTTAA 318
||| ||||| :
Db 2518 GAAATATTTGAG 2577
QY 319 GATCAGGTAATGCGGACGATGACGAGTCAAGTGAAGTGAAGAGAGAGAGAG 371
||| ||||| :
Db 2578 GAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2630

RESULT 3

US-09-357-014-8
; Sequence 8, Application US/09357014
; Patent No. 6291645

GENERAL INFORMATION:
APPLICANT: Jaekyoon Shin, Insil Jouny, Ratna K. Vadlamudi
and Jack L. Strominger
TITLE OF INVENTION: P62 POLYPEPTIDES, RELATED POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/357,014
FILING DATE: 19-Jul-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/574,959
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 439..3157
US-09-357-014-8

Query Match 10.7%; Score 40.2; DB 4; Length 3211;
Best Local Similarity 52.0%; Pred. No. 0.0066;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 199 GAACACGCGTCAACAGAAATGTTACTACATCTCTGATGTTGAGAGTTAGACCA 258
||||| ||||| :
Db 2458 GAACAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2517
QY 259 GAAATATTTGTTGCGAGAAACTTCAAGAAATATTTGACAGCTCGAGGTTAA 318

TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSK/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match
Best Local Similarity 47.2%; Score 39.6; DB 2; Length 3489;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAGCTCTGATCTTGACGAAATGCAAAAAACATGTTTCAAGAG 177
DB 871 GAGAAGACATCCAGATGAAAAAAGATGTCGAAATATATGAGCTGCGAGATTAAT 930
QY 178 GAGTACGCTACATCGTTTGAAGACACGCTCAACAGGAATGTTACTACATGTCGTT 237
DB 931 GGGGACACAGAGATTAGCAAGAAAGTGCAGTGCAGAGATGACATGATTAAGGAT 990
QY 238 GATGTTGAAGATTAGACCAAGAAAAATTTGTTGCGACGAAACTTCACGGAATTTAT 297
DB 991 GATGAGGAGGACGAGAGACAGATGAGAGGACGAGAGAGATGACGAGAGATGACGAG 1050
QY 298 TTGACAGACTGCGAGGTTAAAGATGCGTAATGCGGAGAGTACAGTGCAGTACGAT 357
DB 1051 GAGGATGACGAGAGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAG 1110
QY 358 GAAGTTGATGAGA 371
DB 1111 GAGGATGACGAGGA 1124

RESULT 7
US-09-298-568-1
Sequence 1, Application US/09298568
Patent No. 6322792
GENERAL INFORMATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kieft, Elliott D.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
FILE REFERENCE: 16412-10001R

CURRENT APPLICATION NUMBER: US/09/298,568
CURRENT FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match
Best Local Similarity 47.2%; Score 39.6; DB 4; Length 3489;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAGCTCTGATCTTGACGAAATGCAAAAAACATGTTTCAAGAG 177
DB 871 GAGAAGACATCCAGATGAAAAAAGATGTCGAAATATATGAGCTGCGAGATTAAT 930
QY 178 GAGTACGCTACATCGTTTGAAGACACGCTCAACAGGAATGTTACTACATGTCGTT 237
DB 931 GGGGACACAGAGATTAGCAAGAAAGTGCAGTGCAGAGATGACATGATTAAGGAT 990
QY 238 GATGTTGAAGATTAGACCAAGAAAAATTTGTTGCGACGAAACTTCACGGAATTTAT 297
DB 991 GATGAGGAGGACGAGAGACAGATGAGAGGACGAGAGAGATGACGAGAGATGACGAG 1050
QY 298 TTGACAGACTGCGAGGTTAAAGATGCGTAATGCGGAGAGTACAGTGCAGTACGAT 357
DB 1051 GAGGATGACGAGAGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAG 1110
QY 358 GAAGTTGATGAGA 371
DB 1111 GAGGATGACGAGGA 1124

RESULT 8
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 10.6%; Score 39.6; DB 2; Length 32207;
Best Local Similarity 47.2%; Pred. No. 0.03;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAATCTTTTAAAGACTCTCTGATCTTGACGAATGCAGAAACATGTTTCAAGACG 177
DB 21126 GAAGAAAGCATCCAGAAATGAAAAAGAAATGTTCCGAAATTAATCAGCTGCGAGAGATAT 21067
QY 178 GAGTACTGCTACATCGTTTGAAGACACGGTCAACAAGAAATGTACTACATGTGCTT 237
DB 21066 GGGGACAAACAGATTAGCAAGAAAGTCAAGATGACATGACATTAAGAT 21007
QY 238 GATGTTGAAGATTAGACCAAGAAATTTGTTGCGACGAAACTTCACGAAATTTAT 297
DB 21006 GATGAGGAGACAGACAGACAGATGAGAGAGAGAGATGACGAGAGATGACGAG 20947
QY 298 TTGACAGACTGCGAGGTAAGATGATGATGCGGAGGTACAGGTGACGAGTACAGAT 357
DB 20946 GAGGATGACAGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
QY 358 GAAGTTGATGAGA 371
DB 20886 GAGGATGACGAGGA 20873

RESULT 9

US-08-757-669A-20/c
Sequence 20, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424

AUTHOR/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match 10.6%; Score 39.6; DB 4; Length 32207;
Best Local Similarity 47.2%; Pred. No. 0.03;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAATCTTTTAAAGACTCTCTGATCTTGACGAATGCAGAAACATGTTTCAAGACG 177
DB 21126 GAAGAAAGCATCCAGAAATGAAAAAGAAATGTTCCGAAATTAATCAGCTGCGAGATAT 21067
QY 178 GAGTACTGCTACATCGTTTGAAGACACGGTCAACAAGAAATGTACTACATGTGCTT 237
DB 21066 GGGGACAAACAGATTAGCAAGAAAGTCAAGATGACATGACATTAAGAT 21007
QY 238 GATGTTGAAGATTAGACCAAGAAATTTGTTGCGACGAAACTTCACGAAATTTAT 297
DB 21006 GATGAGGAGACAGACAGACAGATGAGAGAGAGAGATGACGAGAGATGACGAG 20947
QY 298 TTGACAGACTGCGAGGTAAGATGATGATGCGGAGGTACAGGTGACGAGTACAGAT 357
DB 20946 GAGGATGACAGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
QY 358 GAAGTTGATGAGA 371
DB 20886 GAGGATGACGAGGA 20873

RESULT 10

US-09-230-371A-20/c
Sequence 20, Application US/09230371A
Patent No. 6348586
GENERAL INFORMATION:

APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 45185-G-PCT-US
CURRENT APPLICATION NUMBER: US/09/230,371A
CURRENT FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: PCT/US97/13346
PRIORITY FILING DATE: 1997-07-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 20

LENGTH: 32207
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match 10.6%; Score 39.6; DB 4; Length 32207;
Best Local Similarity 47.2%; Pred. No. 0.03;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAATCTTTTAAAGACTCTCTGATCTTGACGAATGCAGAAACATGTTTCAAGACG 177
DB 21126 GAAGAAAGCATCCAGAAATGAAAAAGAAATGTTCCGAAATTAATCAGCTGCGAGATAT 21067
QY 178 GAGTACTGCTACATCGTTTGAAGACACGGTCAACAAGAAATGTACTACATGTGCTT 237
DB 21066 GGGGACAAACAGATTAGCAAGAAAGTCAAGATGACATGACATTAAGAT 21007
QY 238 GATGTTGAAGATTAGACCAAGAAATTTGTTGCGACGAAACTTCACGAAATTTAT 297
DB 21006 GATGAGGAGACAGACAGACAGATGAGAGAGAGATGACGAGAGATGACGAG 20947
QY 298 TTGACAGACTGCGAGGTAAGATGATGATGCGGAGGTACAGGTGACGAGATGACGAG 20887

Db 20946 GAGGATGACGAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
QY 358 GAAGTGTGAAGA 371
Db 20886 GAGGATGACGAGGA 20873

RESULT 11

US-08-766-738-2

; Sequence 2, Application US/08766738

; Patent No. 5916749

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Goll, Surya K.

; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/766,738

; FILING DATE: Herewith

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0177 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 966 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: Consensus

; CLONE: 1813361

; US-08-766-738-2

Query Match 9.6%; Score 36; DB 2; Length 966;
Best Local Similarity 49.5%; Pred. No. 0.077;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACACGCTCAACAAGATGTTACTACAAATGTCGTGATGGTGAAGATTA 252
Db 583 GTTGATNNAAGANGAGAGGAGGAGAGAGATGAGAGAGCGAGCGATGAGAGAT 642
QY 253 GACCAAGAAAATTTGTTGTCGACGAAACTTCAGGAAATTTATTTGACAGCTGCGAG 312
Db 643 GGTGAAGAGAGAGAGGTTTGTGATGAGAGAGATGATGAAGATGTGAAGAGGAGATGAG 702
QY 313 GGTAAAGATGACGATTAATGCGGACGATGACAGAGTGAAGTGAAGTGAAGAT 372
Db 703 GACGACGATGAAGTCACTGAGGAGGAGAGAAATTTGACTTGAAGAGATGAAGAT 762
QY 373 TA 374

Db 763 GA 764

RESULT 12

US-09-262-610-2

; Sequence 2, Application US/09262610

; Patent No. 6428949

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Goll, Surya K.

; TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/262,610

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/766,738

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0177 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 966 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: Consensus

; CLONE: 1813361

; US-09-262-610-2

Query Match 9.6%; Score 36; DB 4; Length 966;
Best Local Similarity 49.5%; Pred. No. 0.077;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACACGCTCAACAAGATGTTACTACAAATGTCGTGATGGTGAAGATTA 252
Db 583 GTTGATNNAAGANGAGAGGAGGAGAGAGATGAGAGAGCGAGCGATGAGAGAT 642
QY 253 GACCAAGAAAATTTGTTGTCGACGAAACTTCAGGAAATTTATTTGACAGCTGCGAG 312
Db 643 GGTGAAGAGAGAGAGGTTTGTGATGAGAGAGATGATGAAGATGTGAAGAGGAGATGAG 702
QY 313 GGTAAAGATGACGATTAATGCGGACGATGACAGAGTGAAGTGAAGTGAAGAT 372
Db 703 GACGACGATGAAGTCACTGAGGAGGAGAGAAATTTGACTTGAAGAGATGAAGAT 762
QY 373 TA 374
Db 763 GA 764

RESULT 13

US-08-867-941-2

Sequence 2, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Qun-Jun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/667,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-2
Query Match
Best Local Similarity 52.2%; Score 35; DB 2; Length 2694;
Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACACGCGTAAACAAGATGTTACTACAT 231
DB 2017 AAAAAGACAAAGGTTATAGCAATATGAAGAAACCATCAAGAAAAAGCCATCAAGAT 2076
QY 232 GTCGTTGATGTTGAAGAGTTAGACCAAGAAATTTGTTGCGACGAACCTTCACGGA 291
DB 2077 TATCTGTTAACGACGACTTCACCCAGAA-----GATGATGACGATGATTTGACCGCA 2130
QY 292 AATTATTGACGACTGCGAGGTTAAAGATGCAAGTAATCGCGCAGTACAGGTGACGAG 351
DB 2131 TCTGATGATTACAAAGATGATGATGCAAGCGGATGATGATTTGATGATTCATCTGATGAT 2190
QY 352 TCAGATGAAGTTGATGAAGT 372
DB 2191 TCACAAAGATGATGACGAGAT 2211
RESULT 14
US-09-074-658-2
Sequence 2, Application US/09074658
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M
APPLICANT: Run-Pan Du
APPLICANT: Qun-Jun Wang
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-795
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-2
Query Match
Best Local Similarity 52.2%; Score 35; DB 4; Length 2694;
Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACACGCGTAAACAAGATGTTACTACAT 231
DB 2017 AAAAAGACAAAGGTTATAGCAATATGAAGAAACCATCAAGAAAAAGCCATCAAGAT 2076
QY 232 GTCGTTGATGTTGAAGAGTTAGACCAAGAAATTTGTTGCGACGAACCTTCACGGA 291
DB 2077 TATCTGTTAACGACGACTTCACCCAGAA-----GATGATGACGATGATTTGACCGCA 2130
QY 292 AATTATTGACGACTGCGAGGTTAAAGATGCAAGTAATCGCGCAGTACAGGTGACGAG 351
DB 2131 TCTGATGATTACAAAGATGATGATGCAAGCGGATGATGATTTGATGATTCATCTGATGAT 2190
QY 352 TCAGATGAAGTTGATGAAGT 372
DB 2191 TCACAAAGATGATGACGAGAT 2211

RESULT 15
US-08-867-941-1
Sequence 1, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loomore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Qun-Jun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-1

Query Match 9.3%; Score 35; DB 2; Length 7650;
Best Local Similarity 52.2%; Pred. No. 0.42;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
QY 172 AAGACGAGTACTGCTACATGTTTGAAGACACAGGTCACACAGAGATGTTACTACAT 231
DB 2157 AAAAAAGACAAAGTTATAGCAATATAGAAACCAAGAAAGGCGCATCAAGAT 2216
QY 232 GTCCGTGATGTTGAAGACTTAGACCAAGAAAATTTGTTGCGACGAAACTTCACGAA 291
DB 2217 TATCTGTTAACCGAGACTTCACCCAGAA-----GATGATGACGATGATTGACCGCA 2270
QY 292 AATTATTGACAGACTGCGAGGTAAGATGCAAGTAATGCGCGCAGGTACAGGTGACGAG 351
DB 2271 TCTGATGATTACAAAGATGATGATGCAATGCGCATGATGATTGATGATGATGAT 2330
QY 352 TCAGATGAGTTGATGAGAT 372
DB 2331 TCACAGATGATGACCGAGAT 2351

Search completed: May 5, 2003, 17:46:42
Job time : 76 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 16:43:12 ; Search time 138 Seconds
(without alignments)
3210.927 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 375
Sequence: 1 atgaattatcttgcatttc.....atgaattatgaattata 375

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues
Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_MA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39.6	10.6	462	9	US-09-918-995-348
2	39.6	10.6	766	10	US-09-864-761-19608
3	39.6	10.6	1944	10	US-09-864-761-2825
4	37.2	9.9	315	10	US-09-864-761-21723
5	36.6	9.8	499	10	US-09-783-590-5790
6	36	9.6	381	9	US-10-015-219-700
7	36	9.6	381	10	US-09-777-564-700
8	36	9.6	966	9	US-10-213-700-2
9	35.8	9.5	204	10	US-09-864-761-21008
10	35.8	9.5	474	10	US-09-864-761-4255
11	35.4	9.4	876	10	US-09-770-445-544
12	35.4	9.4	1092	9	US-09-938-842A-1125
13	35	9.3	267	10	US-09-878-574-15076
14	35	9.3	520	9	US-10-184-644-332
15	35	9.3	520	10	US-10-184-634-332
16	35	9.3	563	10	US-09-864-761-13293
17	34.8	9.3	327	10	US-09-864-761-13293
18	34.8	9.3	1626	9	US-10-116-016-19
19	34.8	9.3	1626	10	US-09-764-848-19

20	34.8	9.3	5361	9	US-09-742-096-2	Sequence 2, Appl1
21	34.8	9.3	6152	9	US-09-742-096-1	Sequence 1, Appl1
22	34.6	9.2	438	10	US-09-864-761-4988	Sequence 4988, Ap
23	34.6	9.2	542	10	US-09-864-761-12239	Sequence 12239, A
24	34.6	9.2	8391	9	US-10-151-736-3	Sequence 3, Appl1
25	34.6	9.2	8493	9	US-10-151-736-5	Sequence 5, Appl1
26	34.6	9.2	1691139	9	US-10-067-514-1	Sequence 1, Appl1
27	34.4	9.2	846	9	US-09-815-242-8918	Sequence 8918, Ap
28	34.4	9.2	1282	9	US-10-002-344A-89	Sequence 89, Appl
29	34.4	9.2	8345	9	US-09-764-691-8451	Sequence 8451, Ap
30	34.2	9.1	381	10	US-09-864-761-21610	Sequence 21610, A
31	34.2	9.1	659158	9	US-09-771-208-20	Sequence 20, Appl
32	33.8	9.0	1824	10	US-09-815-242-9281	Sequence 9281, Ap
33	33.8	9.0	1824	10	US-09-815-242-9281	Sequence 9281, Ap
34	33.6	9.0	305	10	US-09-864-761-19262	Sequence 19262, Ap
35	33.6	9.0	496	10	US-09-864-761-2534	Sequence 2534, Ap
36	33.4	8.9	420	10	US-09-864-761-23266	Sequence 23266, A
37	33.4	8.9	451	10	US-09-864-761-6554	Sequence 6554, Ap
38	33.4	8.9	451	10	US-09-864-761-14152	Sequence 14152, A
39	33.4	8.9	7657	10	US-09-070-927A-33	Sequence 33, Appl
40	33.2	8.9	157	10	US-09-864-761-29856	Sequence 29856, A
41	33.2	8.9	423	9	US-09-938-842A-592	Sequence 592, App
42	33.2	8.9	58985	9	US-09-901-152-3	Sequence 3, Appl1
43	33	8.8	432	7	US-08-781-986A-696	Sequence 696, App
44	33	8.8	1959	10	US-09-864-761-4012	Sequence 4012, Ap
45	33	8.8	2920	10	US-09-801-574-87	Sequence 87, Appl

ALIGNMENTS

RESULT 1
US-09-918-995-348
; Sequence 348, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-736
; CURRENT APPLICATION NUMBER: US/09/918, 995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 348
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)..(462)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-348

Query Match 10.6%; Score 39.6; DB 9; Length 462;
Best Local Similarity 50.5%; Pred. No. 0.11;
Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY	186	CTACATCGTTTAAAGCAGCGTCAACAGGATGTGTACACAAATCGTGTGATGGA	245
DB	164	CATACAGCAGTATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATGA	223
QY	246	AGAGTTAGACCAAGAAAATTTTGTTCGACGAAACCTTACGGAATTTATTCACAGA	305
DB	224	AGAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTTGAGA	283
QY	306	CTGCGAGGCTTAATATCGGACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTGA	365
DB	284	GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTGA	343
QY	366	TGAAGATTAA 375	

LENGTH: 1944
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AI08720.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825

Query Match 10.6%; Score 39.6; DB 10; Length 1944;
Best Local Similarity 51.1%; Pred. No. 0.22; Mismatches 89; Indels 0; Gaps 0;
Matches 93; Conservative 0;

QY 187 TACATCGTTTGAAGACACGCTCAACAGGATGTTACTACATGCTGTGATGTCGAA 246
DB 924 TCACACTTGAGACACCATGAGATGATGATGATGATGATGATGATGATGATGAT 865
QY 247 GAGTTAGACAGAAATTTGTTGTCACGAAACTTCACGGAATTTATTTGACAGAC 306
DB 864 GATGAGAGGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 805
QY 307 TCCGAGGGTAAAGATGACAGTATGCGGACGATGACAGGTCAGATGATGATGAT 366
DB 804 GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 745
QY 367 GA 368
DB 744 GA 743

RESULT 4

US-09-864-761-21723
Sequence 21723, Application US/09864761
Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmlca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21723
LENGTH: 315
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AI033533.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: U28921.1, EVALU6 6.506-01
US-09-864-761-21723

Query Match 9.9%; Score 37.2; DB 10; Length 315;
Best Local Similarity 50.0%; Pred. No. 0.43; Mismatches 93; Indels 0; Gaps 0;
Matches 93; Conservative 0;

QY 190 ATCGTTTGAAGACACGCTCAACAGGATGTTACTACATGCTGTGATGTCGAAAG 249
DB 52 ACCGAGTTTGAGAAACAAACAAAAAGAAAGAAAAATCCAGACAGACAGAAACAA 111
QY 250 TTGACCAAGAAATTTGTTGTCACGAAACTTCACGGAATTTATTTGACAGACTGC 309
DB 112 GAACAAAGAAAGAGAAAGAGGCGGAGAAAGAAAGAGAAAGAGAAAGAAAGAA 171
QY 310 GAGGCTAAAGATGAGTATGCGGACGATACAGTGCAGACATGATGATGATGAA 369
DB 172 GAAGAAAGAA 231
QY 370 GATTAA 375
DB 232 GAAGAA 237

RESULT 5

US-09-783-590-5790
Sequence 5790, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0

```
SEQ ID NO 5790
LENGTH: 499
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (9)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (127)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (211)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (233)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (262)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (270)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (302)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (320)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (330)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (337)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (345)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (346)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (348)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (368)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (408)
```

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (417)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (423)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (439)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (456)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (467)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (468)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (469)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (470)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (476)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (481)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (492)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (497)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5790
```

Query Match 9.8%; Score 36.6; DB 10; Length 499;
Best Local Similarity 50.6%; Pred. No. 0.82;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
QY 31 CTGCGAAGCTTGATGCTCACTACTCTTCAGAGAAGCTGAAGATGTTGGACGTT 90
DB 10 CTAGATTTCTAGTAATCTAAACCCCTCCAGGTAGCCAAAGATCAAGTTGTTGAC 69
QY 91 TACGGACAGCAAAATATATACAGACTTCGATTAACCTTTAGAGAGTCCCTGATCTGAC 150
DB 70 TCCACAGCACACAAAGAGATGCTCTAGATTAAGACAGTAAGAGATTTATATCATATAT 129
QY 151 GAATGCAAAAACATGTTTCAAGACGAGTACTGCTACATCGTTTGAAG 202
DB 130 CATTCATGTAATTAACCTTTAATTGAACATTTGAACATGCTGTGAGG 181
```

RESULT 6
US-10-015-219-700
Sequence 700, Application US/10015219
Publication No. US20030008299A1
GENERAL INFORMATION:
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
OF TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.493C1
CURRENT APPLICATION NUMBER: US/10/015.219
CURRENT FILING DATE: 2002-03-02

NUMBER OF SEQ ID NOS: 1739
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 700
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 66, 201, 213, 225, 251, 255, 261, 262, 265, 280, 290, 303,
LOCATION: 313, 322, 339, 373
OTHER INFORMATION: n = A,T,C or G
US-10-015-219-700

Query Match 9.6%; Score 36; DB 9; Length 381;
Best Local Similarity 49.5%; Pred. No. 1.1;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACACGTCACAGAAATGTTACTACATGCTGATGTTGAGAGCTTA 252
DB 21 GTGATGAAGAT 80
QY 253 GACCAAGAAAATTTGTTGTCGACGAAACTTCACGAAAATTTTTCACAGACTGGAG 312
DB 81 GGTGAAGAT 140
QY 313 GGTAAAGATGACGTAATGCGCAGGTACAGGTGACGAGTCAGATGAAGTGAAGAT 372
DB 141 GACGACGATGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 200
QY 373 TA 374
DB 201 NA 202

RESULT 7
US-09-777-564-700
Sequence 700, Application US/09777564
Patent No. US2002022591A1
GENERAL INFORMATION:
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.493
CURRENT APPLICATION NUMBER: US/09/777,564
NUMBER OF SEQ ID NOS: 1730
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 700
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(381)
OTHER INFORMATION: n = A,T,C or G
US-09-777-564-700

Query Match 9.6%; Score 36; DB 10; Length 381;
Best Local Similarity 49.5%; Pred. No. 1.1;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACGTCACAGAAATGTTACTACATGCTGATGTTGAGAGCTTA 252
DB 21 GTGATGAAGAT 80
QY 253 GACCAAGAAAATTTGTTGTCGACGAAACTTCACGAAAATTTTTCACAGACTGGAG 312
DB 81 GGTGAAGAT 140
QY 313 GGTAAAGATGACGTAATGCGCAGGTACAGGTGACGAGTCAGATGAAGTGAAGAT 372
DB 141 GACGACGATGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAT 200

QY 373 TA 374
DB 201 NA 202

RESULT 8
US-10-213-700-2
Sequence 2, Application US/10213700
Publication No. US2003002232A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/213,700
FILING DATE: 06-Aug-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/766,738
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0177 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: 1813361
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-213-700-2

Query Match 9.6%; Score 36; DB 9; Length 966;
Best Local Similarity 49.5%; Pred. No. 1.7;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAAGACGTCACAGAAATGTTACTACATGCTGATGTTGAGAGCTTA 252
DB 583 GTTATGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 642
QY 253 GACCAAGAAAATTTGTTGTCGACGAAACTTCACGAAAATTTTTCACAGACTGGAG 312
DB 643 GGTGAAGAT 702
QY 313 GGTAAAGATGACGTAATGCGCAGGTACAGGTGACGAGTCAGATGAAGTGAAGAT 372
DB 703 GACGACGATGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAT 762
QY 373 TA 374
DB 763 GA 764

```

RESULT 9
US-09-864-761-21008/c
: Sequence 21008, Application us/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Acomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
: SEQ ID NO 21008
: LENGTH: 204
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL121580.6
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.97
: OTHER INFORMATION: NT_HIT: APL55827.1, EVALUE 1.60e+00
: OTHER INFORMATION: EST_HUMAN HIT: BB889896.1, EVALUE 2.80e+00
US-09-864-761-21008

```

```

Query Match          9.5%; Score 35.8; DB 10; Length 204;
Best Local Similarity 54.1%; Pred. No. 0.88;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 238 GATGCTGAAGATAGACCAAGAAAAATTTGTTGTCAGCAAGAAACTTCACGGAATTTAT 297
      ||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 171 GATGATGAAGGTGATGAAAAATGCAAGATGAAATGTAGAGACGAAGAAGATGATTATGACAA 112

QY 298 TTGACAGACTCGAGCGGTGAAGATGCAGCTATCGCGAGGTACAGAGTACAGACTCATGAT 357
      ||| ||| || || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 111 GATGAAGAAAGATGAAAAATGAAGATGAAAAATGAAGATGAAGATGAAGATGAAGAAAGAA 52

QY 358 GAAGTTGATGAAGAT 372
      || | ||||| |||
Db 51 GATGAATGAAGAT 37

RESULT 10
US-09-864-761-4255/c
; Sequence 4255, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4255
; LENGTH: 474
; TYPE: DNA

```

Query Match	9.48;	Score 35.4;	DB 10;	Length 876;
Best Local Similarity	52.38;	Pred. No. 2.4;		

```

RESULT 13
US-09-878-574-15076
; Sequence 15076, Application US/09878574
; Patent No. US20020110548A1
GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535

```

[illegible]

Query Match	9.3%	Score 35;	DB 9;	Length 520;
Best Local Similarity	24.3%	Pred. No. 2.4;		
Matches 42;	Conservative 35;	Mismatches 96;	Indels 0;	Gaps 0;

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 16:22:32 ; Search time 1576 Seconds
(without alignments)
3853.622 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagtattcttgatc.....atgaagtgcgaagattra 375

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estum:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_frod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.8	13.8	879	17	AZ550718 ENT56718
2	51.6	13.8	701	13	BM168486 ENT571009
3	51.6	13.8	721	13	BM165474 ENT567997
4	51.2	13.7	881	17	AZ688514 ENTMD22TF
5	50	13.3	236	13	BM163498 ENT566021
6	50	13.3	272	13	BM16089 ENT568612

Result No.	Score	Query Match	Length	ID	Description
7	50	13.3	274	13	BM166886 ENT569409
8	50	13.3	276	13	BM163293 ENT568816
9	50	13.3	425	13	BM162507 ENT565030
10	50	13.3	453	13	BM167376 ENT569899
11	50	13.3	459	13	BM171330 ENT571853
12	50	13.3	580	13	BM161175 ENT563698
13	50	13.3	648	13	BM164042 ENT565655
14	50	13.3	651	13	BM166448 ENT568971
15	50	13.3	667	13	BM161293 ENT568816
16	50	13.3	678	13	BM170834 ENT573357
17	50	13.3	702	13	BM169291 ENT571814
18	50	13.3	719	13	BG602689 ENT501779
19	50	13.3	739	13	BM160442 ENT562965
20	50	13.3	753	13	BM167115 ENT569638
21	50	13.3	762	12	BE846498 PSR 206 L
22	50	13.3	790	13	BM166415 ENT568938
23	49.2	13.1	942	17	BM148582 ENTCP82TF
24	49	13.1	816	17	AZ535744 ENTQ25TR
25	49	13.1	891	17	AZ683582 ENTMRK47TR
26	49	13.1	976	17	BM149983 ENTOD93TF
27	48.4	12.9	435	13	BM169710 ENT572233
28	47.4	12.6	900	17	AZ549980 ENTDB44TR
29	47	12.5	843	17	AZ551618 ENTDB54TR
30	47	12.5	877	17	AZ531291 ENTBQ34TR
31	47	12.5	908	17	AZ548467 ENTBK30TR
32	47	12.5	912	17	AZ551092 ENTFLJ22TF
33	47	12.5	931	17	BM160272 ENTQV49TR
34	46.8	12.5	724	13	BM161274 ENT563797
35	46.8	12.5	1006	17	AZ679713 ENT1H56TR
36	46.6	12.4	558	13	BM163267 ENT565790
37	46.6	12.4	890	17	BM146886 ENTMRK48TF
38	46.6	12.4	906	17	BM153606 ENTMS83TF
39	46.4	12.4	468	12	BG553005 ddb82809.
40	46.4	12.4	880	17	AZ669474 ENT188TR
41	46.2	12.3	493	13	BM170539 ENT573082
42	46.2	12.3	1101	17	AL108773 Drosophila
43	45.8	12.2	849	17	AZ546009 ENT5653TF
44	45.8	12.2	905	17	AZ550256 ENT5V58TR
45	45.4	12.1	916	17	AZ671886 ENTMR44TR

ALIGNMENTS

RESULT 1
LOCUS AZ550718/c 879 bp DNA linear GSS 14-NOV-2000
DEFINITION ENT56718 Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ550718
VERSION AZ550718.1 GI:11176019
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 879)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3343
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 25
High quality sequence stop: 801.

FEATURES
source

Location/Qualifiers
1. .879
/organism="Entamoeba histolytica"
/strain="HMI:1MS"
/db_xref="taxon:3759"
/clone_lib="Entamoeba histolytica sheared DNA"
/note="Vector: pHO51; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 193 a 219 c 72 g 395 t
ORIGIN

Query Match 13.8%; Score 51.8; DB 17; Length 879;
Best Local Similarity 46.9%; Pred. No. 0.0004;
Matches 161; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 29 GCGTCGCAAGCTTGTCTGATCTCACTTCTTCAGAGAACGCGAATGTTGAGCGT 88
DB 768 GACTCTGTACTCTGCTTAATGCTTGTCCAGAGAAAGAACTTACACACAT 709
QY 89 TTTACGCGAAGAAATATACAGACTTCGATTAATCTTTTAAAGAGCTCTGATCTTG 148
DB 708 TCGATATTGATTCTCAGTAGTAGTAGAGCTTTTGAAGAAATATATGATG 649
QY 149 ACGAATGCAAAAAACATGTTTCAAGACGAGTACTGCTACATGTTTGAACACGCG 208
DB 648 AAGATGATGAAATTCATAGCAGAAAGAAATGATGAAGATGAAAGATGAG 589
QY 209 TCACACAGAAATGTTACTACATGTCGTTGATGAGTGTGACCAAGAAATTTG 268
DB 588 AAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATG 529
QY 269 TTGTGACGAAAACTTACGCGAAATTTTTCAGACTCGAGGCTAAGATCAGGTA 328
DB 528 AAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATG 469
QY 329 ATGCCGACGTCACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 371
DB 468 ATGAAGAAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATG 426

RESULT 2

LOCUS BM168486 701 bp mRNA linear EST 04-DEC-2001

DEFINITION EST571009 PYBS Plasmodium yoelii yoelii cDNA clone pYCPB36 5' end,
mRNA sequence.

ACCESSION BM168486
VERSION BM168486.1 GI:17301718
KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 701)
AUTHORS Fraser, C.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Carlton, J.M., and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.
Location/Qualifiers
1. .701
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCPB36"
/clone_lib="PYBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BAB6/cby mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 3-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT 320 a 62 c 147 g 172 t
ORIGIN

Query Match 13.8%; Score 51.6; DB 13; Length 701;
Best Local Similarity 55.6%; Pred. No. 0.0004;
Matches 99; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 198 TGAAGACAGCGTCACAGAAATGTTACTACATGTCGTTGATGTCGAGATGACCA 257
DB 512 TGAAGATGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATG 571
QY 258 AGAAAAATTTGTTGACGAAAACTTACGCGAAATTTTTCAGACTCGAGGCTA 317
DB 572 GATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATG 631
QY 318 AGATGCGAATGATGCGGACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTC 375
DB 632 AGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATGAAAGATG 689

RESULT 3

LOCUS BM165474 721 bp mRNA linear EST 04-DEC-2001

DEFINITION EST567997 PYBS Plasmodium yoelii yoelii cDNA clone pYCMN8 5' end,
mRNA sequence.

ACCESSION BM165474
VERSION BM165474.1 GI:17311155
KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 721)
AUTHORS Fraser, C.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Carlton, J.M., and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@ligr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

Location/Qualifiers
1..236

/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="pYCF94"
/clone_1lb="PyBS"
/dev_stage="asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridAP vector and plasmid DNA
isolated."

BASE COUNT 121 a 8 c 68 g 39 t
ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 236;
Best Local Similarity 55.1%; Pred. No. 0.001; Mismatches 80; Indels 0; Gaps 0;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACAGGTCACACAGAAATGTTACTACATGCTGTGATGATGAGAGTTAGACCA 257
||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 53 TGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 112
QY 258 AGAAAAATTTGTTGTCACAGCAAACTTCACGGAATTTATTGACAGACTGCGAGGGTAA 317
- - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 113 GGATGAAGATGACGACGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 172
QY 318 AGATGACAGTATGCGGACGATGACGATGACGATGACGATGACGATGACGATGACGATG 375
||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 173 AGATGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 230

RESULT 6
BM166089 272 bp mRNA linear EST 04-DEC-2001
LOCUS BM166089
DEFINITION mRNA sequence.
ACCESSION BM166089
VERSION BM166089.1 GI:17299321
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Garucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@ligr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

Location/Qualifiers
1..272

/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="pYCN52"
/clone_1lb="PyBS"
/dev_stage="asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridAP vector and plasmid DNA
isolated."

BASE COUNT 137 a 11 c 75 g 49 t
ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 272;
Best Local Similarity 55.1%; Pred. No. 0.0011; Mismatches 80; Indels 0; Gaps 0;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACAGGTCACACAGAAATGTTACTACATGCTGTGATGATGAGAGTTAGACCA 257
||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 89 TGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 148
QY 258 AGAAAAATTTGTTGTCACAGCAAACTTCACGGAATTTATTGACAGACTGCGAGGGTAA 317
- - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 149 GGATGAAGATGACGACGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 208
QY 318 AGATGACAGTATGCGGACGATGACGATGACGATGACGATGACGATGACGATGACGATG 375
||||| - - - - - ||||| - - - - - ||||| - - - - - ||||| - - - - - |||||
Db 209 AGATGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 266

RESULT 7
BM166886 274 bp mRNA linear EST 04-DEC-2001
LOCUS BM166886
DEFINITION EST569409 PyBS Plasmodium yoelii yoelii cDNA clone pYCN214 5' end,
mRNA sequence.
ACCESSION BM166886
VERSION BM166886.1 GI:17300118
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Garucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..276

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCN214"

/clone_1lb="PYBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT
ORIGIN

138 a 11 c 76 g 49 t

Query Match

13.3%; Score 50; DB 13; Length 274;

Best Local Similarity 55.1%; Pred. No. 0.0011;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAACACCGTCACAGCAAGATGTACTACATGTCGTTAGTGAAGTTAGACCA 257

DB 91 TGAAGATGATGAAGAGATGAAGATGATGACGACAGATGATGAAGA 150

QY 258 AGAAAAATTTGTTGTCGACGAAACCTTCACGAAATATTATTTGACAGCTCGAGGGTAA 317

DB 151 GGATGAAGATGACGAGAGAGATGAAGATGACGAGAGATGAAGATGA 210

QY 318 AGATGACAGTAAATGCGGAGGTACAGGTACAGTCAAGTGAAGTTGATGAAGTTAA 375

DB 211 AGATGAAGATGATGAAGAGATGAAGATGATGAAGAGATGAAGAGATGA 268

RESULT 8

LOCUS

BM163293 276 bp mRNA linear EST 04-DEC-2001

DEFINITION EST565816 PYBS Plasmodium yoelii yoelii cDNA clone pYCLD34 5' end,

mRNA sequence.

ACCESSION

BM163293 GI:17308974

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Plasmodium yoelii yoelii.
Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 276)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..276

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCLD34"

/clone_1lb="PYBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT
ORIGIN

139 a 12 c 76 g 49 t

Query Match

13.3%; Score 50; DB 13; Length 276;

Best Local Similarity 55.1%; Pred. No. 0.0011;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAACACCGTCACAGCAAGATGTACTACATGTCGTTAGTGAAGTTAGACCA 257

DB 93 TGAAGATGATGAAGAGATGAAGATGATGACGAGAGATGAAGATGATGAAGA 152

QY 258 AGAAAAATTTGTTGTCGACGAAACCTTCACGAAATATTATTTGACAGCTCGAGGGTAA 317

DB 153 GGATGAAGATGACGAGAGAGATGAAGATGACGAGAGATGAAGATGA 212

QY 318 AGATGACAGTAAATGCGGAGGTACAGGTACAGTCAAGTGAAGTTGATGAAGTTAA 375

DB 213 AGATGAAGATGATGAAGAGATGAAGATGATGAAGAGATGAAGAGATGA 270

RESULT 9

LOCUS

BM162507 425 bp mRNA linear EST 04-DEC-2001

DEFINITION EST565030 PYBS Plasmodium yoelii yoelii cDNA clone pYCLT48 5' end,

mRNA sequence.

ACCESSION

BM162507 GI:17308188

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Plasmodium yoelii yoelii.
Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 425)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR

Unpublished (2001)
Contact: Jane Carlton

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
<http://www.malaria.m4.org/mr4pages/index.html>
Seq primer: ADP.

Location/Qualifiers
1. .425

/organism="Plasmodium yoelli yoelli"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCKT18"
 /clone_11b="PYBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pAD-Gal4: At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dGTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was coprecipitated and ligated to HygriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-Gal4) was excised from the HygriZAP vector and plasmid DNA isolated."

BASE COUNT	202 a	27 c	109 g	87 t
ORIGIN				

Query Match	13.3%	Score 50;	DB 13;	Length 425;
Best Local Similarity	55.1%	Pred. No. 0.0011;		
Matches	98;	Conservative	0;	Mismatches 80;
			Indels	0;
			Gaps	0;

Dy 198 TGAAGACACGGTCAACAAGGATGTCTACTCAAATGTCTTGATCGTGAAGA GTTAAGCCA 257
| | | | |
Db 242 TGAAGATGATGAAGAGATGAAAGATGAAAGATGATGACAGAAGAATGCAAGATGATGAAGA 301

Qy 258 AGAAAAATTTGTTGTCTGCAGCAGAAACCTTCACGGAAAATTATTTGACAGACTCGGAGGGTAA 317
|| || || || || || || || || || || || || || || ||
Db 302 GGATGAAGATGACGAGAAGATGAAGATGACGAGAAGATGAAGATGACGAGAAGATGA 361

```

QY      318 AGATGCAAGGTAAATGCGGCAGGTTACAGTGCACAGTCAGATGAAGTTGTAAGAATTAA 375
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      362 AGATGCAAGTGATGACAGAAAGATGAAAGATGATGTAAGAGAGATGAAGAAGATGAAGATGAA 419

```

RESULT 10	
BM167376	
LOCUS	BM167376
453 bp	mpna
1400000	EST 04-DEC-2001

DEFINITION	ES13b39899 PYBS Plasmodium yoelli yoelli cDNA clone PYC0128 5' end,
ACCESSION	BM167376
VERSION	BM167376.1 GI:17300608

AEIWORDS	SOURCE	ORGANISM
Est.	Plasmodium	yeelll yeelll.
	Plasmodium	yeelll yeelll
	Eukaryota: Alveolata:	Apicomplexa: Haemosporida: Plasmodium

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 453)	Carlton, J. M., Daly, T. M., Long, C. A., Bergman, L. W., Vaidya, A. B., Fraser, C. M., and Carucci, D. J.	Plasmidium yoeili EST protect at TRG

Unpublished (2001)
Contact: Jane Carlton

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mt4.org/mr4pages/index.html>
Seq primer: ADP.

Location/Qualifiers
1. .453

/organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYC0128"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pMD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcytalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HygriZap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pMD-GAL4) was excised from the HygriZap vector and plasmid DNA isolated."

BASE COUNT	223 a	25 c	112 g	93 t
ORIGIN				

Query Match	13.3%	Score 50;	DB 13;	Length 453;
Best Local Similarity	55.1%;	Pred. No. 0.0012;		
Matches 98;	Conservative 0;	Mismatches 80;	Indels 0;	Gaps 0;

QY 198 TGAAGACACGGTCAACAAGGATGTCTACTAATGTCTTGATGCGAAG6TTAGACC A 257
| | | | |
DB 154 TGAAGATCATGAAGAAGATGAAGATGATGACGAAGAAGATGAAGATGATGAAGA 213

OY 258 AGAAAAATTGTTGTTCAGCAGAAACCTTACGGAAAATTTATTGCACAGACTCGGAGGTA 317
 || | | | | | | | | | |
Dδ 214 GGATGAAGATGACGAAGAAGATGAAGATGACGAAGAAGATGA 273

Dy 318 AGATGAGGTAAATCGGCAGCTACAGTGACAGTCAGATGAAGTTGATGAAGATTAA 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 274 AGATGAGATGATGAGAAGAGATGMAAGTGTATGTAAGAGAGATGAAGATGAAGAAGATGAA 331

RESULT 11	
BM171330	
LOCUS	
BM171330	
459 bp	
EBNA	
linear	
EST 04-DEC-2001	

DEFINITION	
ESTS/3853 PYBS Plasmodium yoellii yoe111 cDNA clone pYCOH1 5' end.	
FEATURES	
mRNA sequence.	
ACCESSION	BM171330
VERSION	BM171330 1 GI..1730A562

KEYWORDS	SOURCE	ORGANISM
ESI.	Plasmodium yoelii yoelii.	
	Plasmodium yoelii yoelii	
	Plasmodium yoelii yoelii	
	Alveolata: Apicomplexa: Haemosporida: Plasmodium.	

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 459)	Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdy, A.B., Fraser, C.M. and Carucci, D.J.	Plasmodium yoelii EST project at TIGR

**JOURNAL
COMMENT**

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ARCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

FEATURES
SOURCE

Location/Qualifiers
1. 459
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73293"
/clone="PYCOP11"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybridAP vector and plasmid DNA isolated."

Query Match	13..3%	Score 50:	DB 13:	length 459:
Best Local Similarity	55.1%	Pred. No.	0.0012:	
Matches	98:	Conservative	0:	Mismatches 80: Indels 0: Gaps 0:
QY 198	TGAAGACGCGTCACACAGAAATGTTACTACAAATGTCGTTGATGGTACAGATTTAGACCA	257		
Db 108	TGAAGATGATGAAGAAGATGAAGATGAAGATGATGACCAACAAAGATGAATGATGAGA	167		
QY 258	AGAAAAATTTGTTGTGCACGAAACTTACGAAAAATTATTTGCACAGCTGCGAGGGTAA	317		
Db 168	GGATGGAAGATGACGAAGAAGATGGAAGATGACGAAGAAGATGAAGATGACGAAGAAGATG	227		
QY 318	AGATGCAGCAATGCGCGCAGGTACAGTGCAGAGTCAGATCAAGTGTATATAAGATTAA	375		
Db 228	AGATGGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGA	285		

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
5712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

FEATURES
source

location/Qualifiers
1. 580
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCKB54"
/clone_1lb="PYBS"
/dev_stage="asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBY mice infected with Pyl7XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven terminals were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

	Query Match	13.3%	Score 50:	DB 13:	length 580:
	Best Local Similarity	55.1%	Pred. No.	0.0012:	
Matches	98:	Conservative	0:	Mismatches 80:	Indels 0:
Gaps					0
QY	198 TGAAGACGCGTCMACAAGAATGTACTACAACTCCGTGATGGTGAGAGTTAGACCA	257 			
Db	397 TGAAGATGATGAGGAAGATGAGATGATGATGATGATGCCAAGAAGTGAAGTGAAGA	456 			
QY	258 AGAAAAATTGTTGTCGCAGAAAACCTCAGGAAATTTATTGACAGACTCGAGGGTAA	317 			
Db	457 GGATGGAAGTAGCAAGAAAGATGGAAGATGACGCAAMAGATMAAGATGACGACGAAGATGA	516 			
QY	318 AGATGCAAGTAATGCCGCGAGTACAGGTACAGCATGCAATGTAATGTAAGTTAA	375 			
Db	517 AGATGAAGATGATGAAGAAGATGAAGAATGATCAAGATGATAAAGATGAAGTGA	574 			

RESULT 12
PM161175

LOCUS	BM161175	580 bp	mRNA	linear	EST 04-DEC-2001
DEFINITION	EST563698 PyBS Plasmodium yoelii yoelii cDNA clone pYCKB4 5' end, mRNA sequence.				
ACCESSION	BM161175				
VERSION	BM161175.1	GI:17306856			
KEYWORDS	EST.				
SOURCE	Plasmodium yoelii yoelii.				
ORGANISM	Plasmodium yoelii yoelii.				
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 580)				
AUTHORS	Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B., Fraser,C.M. and Carucci,D.J.				
TITLE	Plasmodium yoelii EST project at TIGR				

RESULT 13

EMBL04042	648 bp	mRNA	linear	EST 04-DEC-2001
LOCUS				
DEFINITION	BM164042	EST565565	PYBS Plasmodium yoelli	cdNA clone
ACCESSION	BM164042			sequence.
VERSION	BM164042.1	GI:17309723		
KEYWORDS	EST.			
SOURCE	Plasmodium yoelli	yoelli.		
ORGANISM	Plasmodium yoelli	yoelli		
REFERENCE	1 (bases 1 to 648)			
AUTHORS	Carlton J.M., Daly T.M., Long C.A., Bergman L.W., Valdivia A.B., Fraser, C.M. and Carucci D.J.			
TITLE	Plasmodium yoelli EST project at TIGR			

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:29:38 ; Search time 56 Seconds
(without alignments)
245.086 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EERDCWTFYANRKYTFDK.....GKDAGNAGTGDSEDEVDED 103

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	573	100.0	103	AA194746	Platelet binding i
2	80.5	14.0	147	AA194746	Platelet binding i
3	80.5	14.0	147	AA194746	Platelet binding i
4	77	13.4	368	AA194746	Platelet binding i
5	75.5	13.2	418	AA194746	Platelet binding i
6	75.5	13.2	419	AA194746	Platelet binding i
7	75.5	13.2	419	AA194746	Platelet binding i
8	75.5	13.2	419	AA194746	Platelet binding i
9	75.5	13.2	432	AA194746	Platelet binding i
10	75.5	13.2	432	AA194746	Platelet binding i

11	73.5	12.8	1154	22	AB166916	Drosophila melanog
12	71.5	12.5	175	21	AA117789	Arabidopsis thalia
13	71.5	12.5	256	21	AA117788	Arabidopsis thalia
14	71.5	12.5	320	21	AA118263	Plasmodium falcipa
15	71	12.4	589	22	AA195086	Human protein sequ
16	70.5	12.3	156	23	AA195086	Lung small cell ca
17	70.5	12.3	367	22	AA195086	Drosophila melanog
18	70	12.2	669	23	AA195086	Amino acid sequenc
19	70	12.2	669	23	AA195086	Amino acid sequenc
20	70	12.2	1167	22	AA195086	Helicobacter pylori
21	69.5	12.1	674	19	AA195086	Streptococcus pneu
22	69.5	12.1	674	23	AA195086	S. pneumoniae Sp11
23	69.5	12.1	3257	22	AA195086	Drosophila melanog
24	69	12.0	717	21	AA195086	Human ORF1583
25	69	12.0	717	22	AA195086	Human polypeptide
26	69	12.0	734	22	AA195086	Human polypeptide
27	69	12.0	734	22	AA195086	Human polypeptide
28	69	12.0	734	22	AA195086	Human polypeptide
29	69	12.0	738	22	AA195086	Human polypeptide
30	69	12.0	738	22	AA195086	Human polypeptide
31	68.5	12.0	610	23	AA195086	Prostate cancer-as
32	68.5	12.0	635	21	AA195086	Plasmodium falcipa
33	68.5	12.0	1869	22	AA195086	Drosophila melanog
34	68	11.9	146	23	AA195086	Murine PLA2 enzyme
35	68	11.9	403	22	AA195086	Human protein sequ
36	68	11.9	561	20	AA195086	zebra mays AAC-11 pr
37	68	11.9	867	22	AA195086	Novel human diagno
38	68	11.9	903	22	AA195086	Human polypeptide
39	68	11.9	961	22	AA195086	Human polypeptide
40	67.5	11.8	117	23	AA195086	Human OREX protein
41	67.5	11.8	366	21	AA195086	Human OREX protein
42	67.5	11.8	450	21	AA195086	Arabidopsis thalia
43	67.5	11.8	450	21	AA195086	Arabidopsis thalia
44	67.5	11.8	488	21	AA195086	Arabidopsis thalia
45	67.5	11.8	488	21	AA195086	Arabidopsis thalia

ALIGNMENTS

RESULT 1
ID AA194746 standard; Protein: 103 AA.
XX AA194746:
AC
DT 29-JAN-2001 (first entry)
XX
DE Platelet binding inhibitor protein Saratin amino acid sequence.
XX
KW Saratin; medicinal leech; thromboembolic disease; intraocular lens;
KW collagen-dependent platelet adhesion inhibition;
KW posterior capsule opacification.
XX
XX Hirudo medicinalis.
OS
XX
PN WO200056885-A1.
XX
XX 28-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-EP02117.
PF
XX
PR 18-MAR-1999; 99EP-010530.
PR 12-MAY-1999; 99EP-0109503.
XX
XX (MERE) MERCK PATENT GMBH.
XX
XX Strittmatter W, Guessow D, Hofmann U, Hemberger J, Fotev Z;
XX Scheudle B;
XX WPI: 2000-611629/58.
XX N-PDB: AAA28180.

PT New Saratin polypeptide and gene isolated from Hirudo medicinalis for
 PT blocking platelet adhesion, especially useful for treating or
 PT preventing thrombotic diseases, or for manufacturing a medicament for
 PT thromboembolic diseases

XX Claim 7; Page 41; 46pp; English.

XX Saratin is a protein isolated from the saliva of the medicinal leech
 CC Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
 CC platelet adhesion. The invention includes polynucleotide sequences
 CC encoding Saratin, an expression vector comprising the DNA sequence, a
 CC host cell transformed with the expression vector, antibodies
 CC immunospecific for Saratin, and methods for identifying Saratin agonists
 CC or antagonists. Saratin is useful for treating thromboembolic diseases,
 CC and for manufacturing a medicament for treating thromboembolic diseases.
 CC It is useful for preventing thrombotic diseases. Saratin is also useful
 CC for coating artificial surfaces, since use of Saratin renders them
 CC non-adhesive for cells and prevents the activation of cells. It may also
 CC be used for coating natural collagen surfaces. Furthermore, Saratin is
 CC useful for modifying intracellular lenses in order to lessen the
 CC thrombogenicity of the lens material, for contacting the lens surface, or
 CC for covalent crosslinking to modify the lens material. The lens material
 CC is used for refractive anterior or posterior chamber ocular implants,
 CC which may be implanted into the eye. This new type of coating avoids
 CC problems contributed by stimulated cell growth. In combination with other
 CC medicaments that are for instance conferring cell death, Saratin coating
 CC helps to completely overcome posterior capsule opacification. The
 CC antibody immunospecific for Saratin, as well as Saratin itself, are
 CC useful for measuring samples derived from host cell cultures or from a
 CC treated subject. The present sequence represents the amino acid sequence
 CC of the Saratin protein.

XX Sequence 103 AA;

Query Match 100.0%; Score 573; DB 21; Length 103;

Best Local Similarity 100.0%; Pred. No. 2.3e-52;

Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEREDCMTFYANKRYTDFDKSFKSSDLDCKKTCFTECYIVFEDTVNKECYVAVDG 60

DB 1 EEREDCMTFYANKRYTDFDKSFKSSDLDCKKTCFTECYIVFEDTVNKECYVAVDG 60

QY 61 EELDOEKVVDENFTENYLTDCGKDGAGNAAGTDESDEVDED 103

DB 61 EELDOEKVVDENFTENYLTDCGKDGAGNAAGTDESDEVDED 103

RESULT 2

AAR22950 ID AAR22950 standard; Protein; 147 AA.

AC AAR22950;

DT 10-SEP-1992 (first entry)

DE Leech antiplatelet protein.

KW LAPP; collagen; platelet aggregation; antithrombotic; anticoagulant;

KW coronary artery disease; cerebrovascular disease; ss.

OS Haementeria officinalis.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Protein /note="21 hydrophobic amino acids"

XX EP480651-A.

PD 15-APR-1992.

PF 07-OCT-1991; 91EP-0309157.

XX 09-OCT-1990; 90US-0594917.

PA (MERI) MERCK & CO INC.

PI Connolly TN, Keller PW;

DR WPT: 1992-125288/16.

DR N-PSDB; AAQ23682.

PT New protein to inhibit collagen stimulated platelet aggregation -

PT extracted from salivary gland of Haementeria officinalis for

PT treating thrombotic diseases

PS Claim 6; Page 15; 20pp; English.

XX When cloned into the yeast expression vector pKHA2, the LAPP gene
 CC sequence forms a fusion product of MAF-alpha-1 leader sequence
 CC plus mature LAPP, which when produced are proteolytically processed by
 CC the Lys-Arg cleaving endopeptidase (KEX2), which cleaves on the
 CC C terminal side of Lys-Arg, and the products secreted into the culture
 CC medium.
 CC LAPP blocks stimulation of platelet aggregation by collagen, at an
 CC antithrombotic concentration of about 100 nMolar (or 1.6 ug/ml).
 CC Platelet aggregation is stimulated by 2 ug/ml collagen and addition
 CC of LAPP inhibits this with an IC50 of 45nM, but this may be
 CC overcome by the addition of 0.25 mM arachidonic acid.
 CC LAPP could be used for treatment and prevention of thrombotic
 CC conditions especially coronary artery and cerebrovascular disease
 CC in mammals.
 CC See also AAQ23871-7.

SQ Sequence 147 AA;

Query Match 14.0%; Score 80.5; DB 13; Length 147;

Best Local Similarity 32.7%; Pred. No. 1.1;

Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

QY 5 DCMTFYANKRYTDFDKSFKSSDLDCKKTCFTEY---CYIVFEDTVNKECYYN 56

DB 66 DCMSKRPQMKLPDNLITKTEFTSVDECKRMCEASVVERSCYIIQINTEFTNCCYRN 120

RESULT 3

AAW68010 ID AAW68010 standard; Protein; 412 AA.

XX AAW68010;

DT 27-APR-1999 (first entry)

DE S. frugiperda immunophilin FRBP46.

KW Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;

KW transplant; tissue graft.

OS Spodoptera frugiperda.

XX Key Location/Qualifiers

FT Domain 111..114

FT /note="EAP motif"

FT Region 124

FT /note="putative nuclear localisation signal"

FT Region 135..138

FT /note="putative nuclear localisation signal"

FT Domain 148..149

FT /note="putative nuclear localisation signal"

FT /note="AP motif"

FT /note="EAP motif"

FT Region 219..223

FT /note="putative nuclear localisation signal"

DR WPI; 2000-464343/40.
 XX
 PT New human Ich-1L and Ich-1S proteins for negative and positive
 PT regulation of programmed cell death and for developing therapeutic
 PT methods for diseases and conditions characterized by cell death, e.g.
 PT myocardial infarction or stroke
 PS
 XX Disclosure; Fig 17; 121pp; English.
 CC The present sequence is a mouse Ice-4 protein sequence. The present
 CC sequence was used in a sequence homology comparison with the protein
 CC sequences of human Ice-4 (Ich-1) (AAB14253), murine
 CC Interleukin-1beta converting enzyme (MICE) (AAB14249), murine ICE2
 CC (AAB14252) and C. elegans ced-3 (AAB14246). The coding sequence of the
 CC present sequence is a member of a family of genes involved in programmed
 CC cell death (apoptosis). Ich-1 may play an important role in both the
 CC positive and negative regulation of apoptosis. The Ich gene may be used
 CC in gene therapy in disorders characterized by cell death e.g. neural and
 CC muscular degenerative diseases, myocardial infarction, stroke, virally
 CC induced cell death and aging.
 CC
 SQ Sequence 418 AA;
 Query Match 13.2%; Score 75.5; DB 21; Length 418;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 QY 44 VFEDTVNKECYNVVDGEEL---DQEKFYVD--ENFTENYL--TDCCKG-DAGNAGTG 94
 DB 26 VFDDLVK-----NVNGDELKIGESASFLNKAKENVLENLEKTDMAKIRAGHIANQ 81
 QY 95 DE-----SDEVD 101
 DB 82 EQLSLQFSNDEDD 94
 RESULT 6
 AAY50109
 ID AAY50109 standard; Protein; 419 AA.
 AC AAY50109;
 XX
 XX 21-JUN-2000 (first entry)
 DE Murine caspase-12 splice variant, caspase-12L.
 XX
 KW Caspase; splice variant; truncated; programmed cell death; apoptosis;
 KW regulation; proteolytic cascade; malignant condition; cancer;
 KW premalignant condition; solid tumour; lymphoma;
 KW chronic lymphocytic leukaemia; prostatic hypertrophy;
 KW preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
 KW ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Cleavage-site 94..95
 FT Protein 95..419
 FT /note="Caspase-12"
 FT Cleavage-site 318..319
 XX
 XX WO952925-A1.
 XX
 XX 21-OCT-1999.
 XX
 XX 14-APR-1999; 99WO-US08064.
 XX
 XX 16-APR-1998; 98OS-0081962.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Yuan J, Morishima N;
 XX
 XX

DR WPI; 1999-620369/53.
 DR N-PSDB; AAZ32645.
 XX
 PT New nucleic acid encoding the short form of caspase-12, used e.g. for
 PT treating tumors
 PS
 XX Claim 1; Fig 1; 68pp; English.
 CC This sequence represents a cDNA encoding murine caspase-12L, a splice
 CC variant of caspase-12. Caspases are a family of proteins involved in
 CC the regulation of apoptosis and are synthesised as proforms which are
 CC activated via cleavage after specific Asp residues. Mammalian cells
 CC express several caspases, and it is thought that these act in a
 CC proteolytic cascade to cause programmed cell death. Nucleic acids
 CC encoding caspase-12S (AAZ32644) or truncated forms of caspase-12L
 CC (AAZ32646, AAZ32647) are used for production, recombinantly or in vivo,
 CC of caspase-12 polypeptides which induce programmed cell death. This is
 CC particularly useful for treating (pre)malignant conditions (e.g., solid
 CC tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
 CC hypertrophy, preneoplastic liver foci and resistance to chemotherapy), or
 CC autoimmune diseases. The caspase-12 proteins can also be used to raise
 CC specific antibodies (for example, to determine gene expression and to
 CC screen expression libraries) or as molecular weight markers. Fragments
 CC of caspase-12 encoding nucleic acids can be used as probes to isolate
 CC the caspase-12 gene (and its allelic variants); in fluorescent in situ
 CC hybridisation for chromosomal location of the caspase-12 gene, and for
 CC Northern blotting to determine caspase-12 mRNA expression in tissues.
 CC
 SQ Sequence 419 AA;
 Query Match 13.2%; Score 75.5; DB 20; Length 419;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 QY 44 VFEDTVNKECYNVVDGEEL---DQEKFYVD--ENFTENYL--TDCCKG-DAGNAGTG 94
 DB 26 VFDDLVK-----NVNGDELKIGESASFLNKAKENVLENLEKTDMAKIRAGHIANQ 81
 QY 95 DE-----SDEVD 101
 DB 82 EQLSLQFSNDEDD 94
 RESULT 7
 AAB14259
 ID AAB14259 standard; Protein; 419 AA.
 AC AAB14259;
 XX
 XX 09-FEB-2001 (first entry)
 DE Mouse Ice-4 protein sequence #1.
 XX
 KW ced-3; virally induced cell death; apoptosis; gene therapy; neural;
 KW muscular degenerative disease; myocardial infarction; stroke; aging;
 KW Interleukin-1beta converting enzyme; ICE; cysteine protease; mouse.
 XX
 OS Mus sp.
 XX
 XX US6083735-A.
 XX
 XX 04-JUL-2000.
 XX
 XX 10-JUN-1994; 94US-0258287.
 XX
 XX 24-JUN-1993; 93US-0080850.
 XX
 XX (GEHO) GEN HOSPITAL CORP.
 XX
 XX Yuan J, Miura M;
 XX
 XX WPI; 2000-464343/40.
 XX N-PSDB; AAZ72843.
 DR

XX New human Ich-1L and Ich-1S proteins for negative and positive
 PT regulation of programmed cell death and for developing therapeutic
 PT methods for diseases and conditions characterized by cell death, e.g.
 PT myocardial infarction or stroke
 PS Disclosure; Fig 16; 121pp; English.
 CC The present sequence is a mouse Ice-4 protein sequence. The present
 CC sequence was used in a sequence homology comparison with the protein
 CC sequences of human Ice-4 (Ich-1) (AAB14253), murine
 CC Interleukin-1beta converting enzyme (mICE) (AAB14249), murine ICE2
 CC (AAB14252) and C. elegans ced-3 (AAB14246). The coding sequence of the
 CC present sequence is a member of a family of genes involved in programmed
 CC cell death (apoptosis). Ich-1 may play an important role in both the
 CC positive and negative regulation of apoptosis. The Ich gene may be used
 CC in gene therapy in disorders characterized by cell death e.g. neural and
 CC muscular degenerative diseases, myocardial infarction, stroke, virally
 CC induced cell death and aging.
 CC
 SQ Sequence 419 AA;
 Query Match 13.2%; Score 75.5; DB 21; Length 419;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 OY 44 VEDIVNRECYNVVDEEL---DQEFVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 26 VEDDLVER---NVLNGDELKIGESAFILNKAEVLENFTLEKTDAGKIFAGHIANSO 81
 OY 95 DE-----SDEVD 101
 DB 82 EQLSLQFSDNEDD 94
 RESULT 8
 AAB29108
 ID AAB29108 standard; protein; 419 AA.
 XX
 AC AAB29108;
 XX
 DT 07-FEB-2001 (first entry)
 XX
 DE Murine caspase-12 protein.
 XX
 KW Mouse; caspase-12; calpain; apoptosis; cell death; autoimmune disease;
 KW cancer; viral infection; apoptotic condition; gene therapy.
 XX
 OS Mus musculus.
 OS
 FN WO200059924-A1.
 XX
 PD 12-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US09173.
 XX
 PR 06-APR-1999; 99US-0127967.
 XX
 PA (HARD) HARVARD COLLEGE.
 XX
 PI Yuan J, Nakagawa T;
 XX
 DR WPI; 2000-679352/66.
 XX
 PT Caspase-12 polypeptide useful for treating apoptotic conditions e.g.
 PT stroke, ischemia and proliferated disease e.g. cancer, lacks at least 5
 PT amino acids in a specific region of the native amino acid sequence -
 XX
 PS Claim 1; Page 65-66; 70pp; English.
 XX
 CC The present invention is concerned with methods for identifying compounds
 CC useful for modulating caspase-mediated apoptosis. These include the
 CC caspase-12 protein shown here. These agents can be used in the treatment

CC of apoptotic conditions, including stroke, neurodegeneration, ischemia,
 CC Alzheimer's disease and muscular dystrophy, and proliferative diseases
 CC such as cancer, autoimmune disorders such as systemic lupus erythematosus
 CC and multiple sclerosis, and viral infections, particularly those
 CC associated with herpes virus, poxvirus and adenovirus. The treatment may
 CC be in the form of gene therapy.
 CC
 SQ Sequence 419 AA;
 Query Match 13.2%; Score 75.5; DB 21; Length 419;
 Best Local Similarity 35.6%; Pred. No. 13;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 OY 44 VEDIVNRECYNVVDEEL---DQEFVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 26 VEDDLVER---NVLNGDELKIGESAFILNKAEVLENFTLEKTDAGKIFAGHIANSO 81
 OY 95 DE-----SDEVD 101
 DB 82 EQLSLQFSDNEDD 94
 RESULT 9
 AAR66769
 ID AAR66769 standard; protein; 432 AA.
 XX
 AC AAR66769;
 XX
 DT 12-SEP-1995 (first entry)
 XX
 DE Human interleukin-1 beta converting enzyme homolog (Ice-4).
 XX
 KW Human interleukin-1 beta converting enzyme homolog; Ice-4;
 KW oncogene bcl-2; programmed cell death; cancer treatment.
 XX
 OS Homo sapiens.
 OS
 PN WO9500160-A.
 XX
 PD 05-JAN-1995.
 XX
 PF 10-JUN-1994; 94WO-US06630.
 XX
 PR 24-JUN-1993; 93US-0080850.
 XX
 PA (GENO) GEN HOSPITAL CORP.
 XX
 PI Miura M, Yuan J;
 XX
 DR WPI; 1995-051742/07.
 XX
 DR N-PSDB; AAO79969.
 XX
 PT Promoting or preventing programmed cell death in vertebrate cells
 PT by inhibiting the activity of interleukin-1 beta converting
 PT enzyme.
 XX
 PS Claim 24; Fig 16; 116pp; English.
 XX
 CC AAO79969 encodes AAR66769 human interleukin-1 beta converting enzyme
 CC homolog (Ice-4), increasing Ice-4s enzymatic activity can
 CC promote the programmed cell death of cancer cells (pref. those
 CC overexpressing the bcl-2 oncogene), this can be used as the basis
 CC of a new cancer treatment. Alternatively by reducing Ice-4s enzymatic
 CC activity programmed cell death can be inhibited, this may be useful
 CC in the development of new cell lines which remain viable in culture
 CC for extended or indefinite periods, independent of growth factors.
 XX
 SQ Sequence 432 AA;
 Query Match 13.2%; Score 75.5; DB 16; Length 432;
 Best Local Similarity 35.6%; Pred. No. 14;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVNKECYNVVDGEEL----DOEKFYVD--ENFTENYL--TDCEGK-DAGNAAGTG 94
 DB 40 VFEDLVK-----NVNGDELLKIGESASFIINKAENLVENFLEKTRDMAGKIFAGHIANSG 95
 QY 95 DE-----SDEVD 101
 DB 96 EQLSLQFSNDEDD 108

RESULT 10

AAR98464
 ID AAR98464 standard; Protein; 432 AA.

AC AAR98464;

DT 25-SEP-1996 (first entry)

DE Murine Ich-3.

KW Ich-3; murine ICE-ced-3 homologue; programmed cell death;
 apoptosis; interleukin-1 beta converting enzyme; gene therapy.

OS Mus sp.

PN W09620721-A1.

PD 11-JUL-1996.

PF 04-JAN-1996; 96MO-US00177.

PR 04-JAN-1995; 95US-0368704.

PA (GEHO) GEN HOSPITAL CORP.

PI Miura M, Yuan J;

DR WPI: 1996-333763/33;

DR N-PSDB; AAT31554.

PT Preventing or promoting programmed cell death in vertebrate cells

PT comprises inhibiting or increasing the activity of

PT Interleukin-1-beta converting enzyme, or altering expression of

PT other related genes

PS Claim 24; Fig 14; 127pp; English.

CC Ich-3 (AAR98464) causes programmed cell death and shows significant

CC homology to mouse Interleukin-1 beta converting enzyme (ICE).

CC mouse Ich-2 (AAR98461) and human Ich-1 (AAR98462-63). Its sequence

CC was deduced from the Ich-3 gene (AAT31554) isolated from a mouse thymus

CC cDNA library. The protein can be obtained from host cells contg.

CC vectors that include an Ich-3 coding sequence. It can be used to

CC control the programmed cell death of vertebrate cells; to develop cell

CC lines that remain viable for extended periods, and to increase the

CC activity of ICE.

CC

QY Sequence 432 AA;

Query Match 13.2%; Score 75.5; DB 17; Length 432;

Best Local Similarity 35.6%; Pred. No. 14;

Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

ID ABB66916 standard; Protein; 1154 AA.

AC ABB66916;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 27540.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL11019.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Disclosure; SEQ ID NO 27540; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (AAB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.

CC

QY Sequence 1154 AA;

Query Match 12.8%; Score 73.5; DB 22; Length 1154;

Best Local Similarity 28.1%; Pred. No. 73;

Matches 27; Conservative 14; Mismatches 42; Indels 13; Gaps 4;

QY 12 NRKYDPEKSPKSSDDDECK-KTQCFKTEYCYIVFEDTVNKECYNVVDGEELDQEKFV 69

DB 505 SKKDDIDFSKFSMDDDKDKLFRVAELMVRCQFLMEKRMQDLGKNKCVNEEKDNVNO 564

QY 70 VDENTENYL--TDCEGRDA-----GNAAGTG 94

DB 565 LVENFWRHYIQCCGCKADKALSNPSMLGNMEYRG 600

RESULT 12

ID AAG17789

AC AAG17789 standard; Protein; 175 AA.

DT AAG17789;

DE 17-OCT-2000 (first entry)

KW Arabidopsis thaliana protein fragment SEQ ID NO: 18943.

Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126284.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134286.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
18-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0138094.
08-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
10-JUN-1999; 99US-0139119.
14-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139889.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
26-JUL-1999; 99US-0145224.
27-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
03-AUG-1999; 99US-0146389.
04-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
05-AUG-1999; 99US-0147192.
06-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
09-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
16-AUG-1999; 99US-0148368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.5%; Score 71.5; DB 21; Length 175;
Best Local Similarity 28.6%; Pred. No. 12;
Matches 26; Conservative 12; Mismatches 32; Indels 21; Gaps 4;

QY 23 KNSSDDECKKTCFKTEYCYIVFEDTVNKECYVNVGDEL-DOEKFYVDENFTENYLT- 80
DB 93 KNGS-----KRALPESEFTWFTDAHKE-----DAGEIHDEVAADIKEDLMSNPPLY 141
QY 81 -----DCEKGAGNAAGTGDSEDEVED 103
DB 142 FNNDADEDEFGDDGDEGEDEDDDEED 172

RESULT 13
AAG17788
ID AAG17788 standard; Protein: 256 AA.

XX AAG17788;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 18942.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR	15-SEP-1999;	99US-0154019;	
PR	16-SEP-1999;	99US-0154039;	
PR	20-SEP-1999;	99US-0154779;	
PR	22-SEP-1999;	99US-0155139;	
PR	23-SEP-1999;	99US-0155486;	
PR	24-SEP-1999;	99US-0155659;	
PR	28-SEP-1999;	99US-0156458;	
PR	29-SEP-1999;	99US-0156596;	
PR	04-OCT-1999;	99US-0157117;	
PR	05-OCT-1999;	99US-0157753;	
PR	06-OCT-1999;	99US-0157863;	
PR	07-OCT-1999;	99US-0158029;	
PR	08-OCT-1999;	99US-0158232;	
PR	12-OCT-1999;	99US-0158369;	
PR	13-OCT-1999;	99US-0159293;	
PR	13-OCT-1999;	99US-0159294;	
PR	13-OCT-1999;	99US-0159295;	
PR	14-OCT-1999;	99US-0159329;	
PR	14-OCT-1999;	99US-0159330;	
PR	14-OCT-1999;	99US-0159331;	
PR	14-OCT-1999;	99US-0159637;	
PR	14-OCT-1999;	99US-0159638;	
PR	18-OCT-1999;	99US-0159584;	
PR	21-OCT-1999;	99US-0160741;	
PR	21-OCT-1999;	99US-0160767;	
PR	21-OCT-1999;	99US-0160768;	
PR	21-OCT-1999;	99US-0160770;	
PR	21-OCT-1999;	99US-0160814;	
PR	21-OCT-1999;	99US-0160815;	
PR	22-OCT-1999;	99US-0160980;	
PR	22-OCT-1999;	99US-0160981;	
PR	22-OCT-1999;	99US-0160989;	
PR	25-OCT-1999;	99US-0161404;	
PR	25-OCT-1999;	99US-0161405;	
PR	25-OCT-1999;	99US-0161406;	
PR	26-OCT-1999;	99US-0161359;	
PR	26-OCT-1999;	99US-0161360;	
PR	26-OCT-1999;	99US-0161361;	
PR	28-OCT-1999;	99US-0161920;	
PR	28-OCT-1999;	99US-0161992;	
PR	28-OCT-1999;	99US-0161993;	
PR	29-OCT-1999;	99US-0162142;	
Query Match	12.5%;	Score 71.5;	DB 21; Length 256;
Best Local Similarity	28.6%;	Pred. No. 19;	
Matches 26; Conservative 12; Mismatches 32; Indels 21; Gaps 4;			
QY	23 KKSLSDECKTKTCFTEKECYIYFEDYVNRKCYVNVDSGEEL-DQKRFVVDENFTENYLT- 80		
Db	174 KKS-----KRALPEESFFTWFTDAQHKR-----DAGDEIHDEVAIDIKEEDLMSNPITY 222		
QY	81 -----DCECKDAGNNAAGTGESDEVEDD 103		
Db	223 FNNDADEEDFDGDDDGDEGEEDDDDEEDD 253		
RESULT 14			
1D	AAB18263		
XX	AAB18263 standard; Protein; 320 AA.		
XX	AAB18263;		
XX			
DT	07-NOV-2000 (first entry)		
XX			
DE	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:130.		
XX			
KW	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;		
KW	antimalarial; malaria; protozoacide; infection; insecticide.		
XX			
OS	Plasmodium falciparum.		
XX			
IN	WO200025728-A2.		
XX			

PD 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1999; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI
XX MPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure: Page 283-284; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the plasmid chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAT0078 to AAT0087 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 320 AA;
SQ
XX
XX Query Match 12.5%; Score 71.5; DB 21; Length 320;
Best Local Similarity 27.8%; Pred. No. 25;
Matches 27; Conservative 9; Mismatches 24; Indels 37; Gaps 5;
OY 9 FYANRKY-----TDFDKS-----FKSSDLECKK-----TCFTEYC----- 41
DB 51 FYALKKRKSLKRRDFTSKSNNDKISIKSYDDFKNELQITIDIKNECYLCEGITTNY 110
OY 42 ---YIVFEDIVNKECYNVVDGSELDOEKFYVDENFT 75
DB 111 DEVYITIEYENDSIL-----KDFEYFVLDKNFT 140

RESULT 15
AAB95086
ID AAB95086 standard; Protein; 589 AA.
XX
XX AAB95086;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:16999.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
PN

XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 16999; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAI03166 to AAI13628 and
CC AAI13633 to AAI18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAI13629 to AAI13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 589 AA;
SQ
XX
XX Query Match 12.4%; Score 71; DB 22; Length 589;
Best Local Similarity 23.8%; Pred. No. 59;
Matches 30; Conservative 23; Mismatches 37; Indels 36; Gaps 6;
OY 12 NRKYTDPRKFRKS-----SDLECKKT-----CFKTE---XCYIYFE 46
DB 158 SKFQKKKKKKKKNIVQHTTDSLEKORTLDGSEIVKPRLECKSTRTREMSVOLI 217
OY 47 DVNKECYNVVDGSELDOEKFYVD-----ENFTENYLTDCGKDAAGAGTGDGS 97
DB 218 MTRDSGTYENSTDGEMCKDALDEDSSEVSIGSESENETTSV-GRASDDDGSEDD 276
OY 98 DEVEDD 103
DB 277 EE-DDD 281

Search completed: May 5, 2003, 17:50:37
Job time : 59 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:45:22 ; Search time 22 Seconds

(without alignments)
137.753 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EBERDCWTFYANRKYTFDK.....GKDNAGTGDSEDEVED 103

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents:AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	80.5	14.0	126	1 US-08-044-547-3	Sequence 3, Appl1
2	80.5	14.0	147	1 US-08-044-547-1	Sequence 1, Appl1
3	80.5	14.0	412	2 US-08-741-134-2	Sequence 2, Appl1
4	75.5	13.2	172	4 US-09-187-789-16	Sequence 16, Appl1
5	75.5	13.2	172	4 US-09-139-600-11	Sequence 11, Appl1
6	75.5	13.2	418	3 US-08-258-287B-58	Sequence 38, Appl1
7	75.5	13.2	418	3 US-08-368-704C-56	Sequence 56, Appl1
8	75.5	13.2	419	3 US-08-258-287B-57	Sequence 57, Appl1
9	75.5	13.2	419	3 US-08-368-704C-55	Sequence 55, Appl1
10	69.5	12.1	124	1 US-08-170-360-1	Sequence 1, Appl1
11	69.5	12.1	674	4 US-08-961-083-200	Sequence 200, App
12	67	11.7	124	1 US-08-170-360-2	Sequence 2, Appl1
13	67	11.7	124	2 US-08-888-497-41	Sequence 41, Appl1
14	67	11.7	124	4 US-09-362-230-41	Sequence 41, Appl1
15	67	11.7	124	5 PCT-US94-07926-41	Sequence 41, Appl1
16	67	11.7	146	2 US-08-888-497-34	Sequence 34, Appl1
17	67	11.7	146	4 US-09-362-230-34	Sequence 34, Appl1
18	67	11.7	146	5 PCT-US94-07926-34	Sequence 34, Appl1
19	67	11.7	299	4 US-09-069-023-20	Sequence 20, Appl1
20	66.5	11.6	86	4 US-09-134-001C-3986	Sequence 3986, Ap
21	66.5	11.6	2182	2 US-08-487-826B-16	Sequence 16, Appl1
22	65.5	11.4	725	2 US-08-813-940-25	Sequence 25, Appl1
23	64.5	11.3	126	1 US-08-170-360-3	Sequence 3, Appl1
24	64.5	11.3	126	2 US-08-888-497-38	Sequence 38, Appl1
25	64.5	11.3	126	4 US-09-362-230-38	Sequence 38, Appl1
26	64.5	11.3	126	5 PCT-US94-07926-38	Sequence 38, Appl1
27	64.5	11.3	148	2 US-08-888-497-36	Sequence 36, Appl1

28	64.5	11.3	148	4 US-09-362-230-36	Sequence 36, Appl1
29	64.5	11.3	148	5 PCT-US94-07926-36	Sequence 36, Appl1
30	64	11.2	519	4 US-09-172-841-55	Sequence 55, Appl1
31	63	11.0	294	4 US-09-355-160D-8	Sequence 8, Appl1
32	63	11.0	1658	2 US-08-609-049A-13	Sequence 13, Appl1
33	63	11.0	1658	4 US-09-170-996-13	Sequence 13, Appl1
34	63	11.0	1726	2 US-08-609-049A-30	Sequence 30, Appl1
35	63	11.0	1726	4 US-09-170-996-30	Sequence 30, Appl1
36	62.5	10.9	250	2 US-08-861-269-5	Sequence 5, Appl1
37	62.5	10.9	250	2 US-09-134-596-5	Sequence 5, Appl1
38	62.5	10.9	250	3 US-09-293-273-5	Sequence 5, Appl1
39	62.5	10.9	1172	1 US-08-313-288B-19	Sequence 19, Appl1
40	62.5	10.9	2555	3 US-08-058-489-36	Sequence 36, Appl1
41	61.5	10.7	346	2 US-08-702-153-2	Sequence 2, Appl1
42	61.5	10.7	346	2 US-08-702-153-4	Sequence 4, Appl1
43	61.5	10.7	416	2 US-09-211-930-3	Sequence 3, Appl1
44	61.5	10.7	416	3 US-09-340-993-3	Sequence 3, Appl1
45	61.5	10.7	416	4 US-09-468-442-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-044-547-3
Sequence 3, Application US/08044547
Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-Stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P. O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-044-547-3

Query Match 14.0%; Score 80.5; DB 1; Length 126;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

TELECOMMUNICATION INFORMATION:

; MOLECULE TYPE: peptide
ITS-08-368-70AC-56

US-08-368-704C-55

CITY: Washington;

ST

```

COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-1

Query Match      12.1%; Score 69.5; DB 1; Length 124;
Best Local Similarity 27.1%; Pred. No. 2.9;
Matches 16; Conservative 8; Mismatches 20; Indels 15; Gaps 1;

QY 28 LDECKKTCFTEYCYIYFEDTVNCKCYNNVDSGELDQEKYFVDENTENYLLDCEKSD 86
Db 38 VDELDRCETHDNCY-----RDANKLDSCKFLVNDPYTESYSCSNTE 81

RESULT 11
US-08-961-083-200
; Sequence 200, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-200

Query Match      12.1%; Score 69.5; DB 4; Length 674;
Best Local Similarity 23.6%; Pred. No. 23;
Matches 25; Conservative 17; Mismatches 35; Indels 29; Gaps 5;

QY 7 WTFYANKYTDPEKSKSDLDCEKT-----CFTEYCYIYFEDTVNCKCYNNVDSG 61
Db 322 WIKOLNKRLEKFD-----SEIINCRTSEQISCYSDISYTF-----LRYFADPV 368

QY 62 ELDOEKF-----VDENFTENYLLDCEG-----KDAGNAAGTGDE 96
Db 369 QEDKALYLDODLVYTKNDDLPATDLDQYPLAAYRFGGRATFGQE 414

RESULT 12
US-08-170-360-2
; Sequence 2, Application US/08170360
; Patent No. 5656602
; GENERAL INFORMATION:
; APPLICANT: Tseng, Albert P. S.
; APPLICANT: Ingllis, Adam
; TITLE OF INVENTION: PLA2 INHIBITORY COMPOUNDS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg Ernst & Kurz
; STREET: Suite 701-E, 535 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/170,360
; FILING DATE: 03-MAR-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PCT/AU92/00333
; FILING DATE: 06-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PK 7058
; FILING DATE: 04-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Ernst, Barbara G.
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1871-104A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-2

Query Match 11.7%; Score 67; DB 1; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKTKCTKTEYCYIYEDTVNKECYNVVGDGEELDOEKFEVDENFTENTYLTDCCEKD 86
DB 40 DLDRG---CQTHDHCY-----NQAKKLESCKFLIDNPYNTYSYKC----- 77

QY 87 AGNAGTGDESDEVD 101
DB 78 SGNVITCSDKNNDC 92

RESULT 13

US-08-888-497-41
Sequence 41, Application US/08888497
Patent No. 5972677

GENERAL INFORMATION:

APPLICANT: Tischnfield, Jay A.

APPLICANT: Sellhammer, Jeffrey J.

TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide

TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences

TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide

TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &

STREET: 200 East Broward Boulevard

CITY: Fort Lauderdale

STATE: FL

COUNTRY: USA

ZIP: 33301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/888,497

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/651,405

FILING DATE:

APPLICATION NUMBER: US 08/097,354

FILING DATE: 26-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Manso, Peter J.

REGISTRATION NUMBER: 32,264

REFERENCE/DOCKET NUMBER: IN21044-5

TELECOMMUNICATION INFORMATION:

TELEPHONE: 305-527-2498

TELEFAX: 305-764-4996

INFORMATION FOR SEQ ID NO: 41:

SEQUENCE CHARACTERISTICS:

LENGTH: 124 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-888-497-41

Query Match 11.7%; Score 67; DB 2; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKTKCTKTEYCYIYEDTVNKECYNVVGDGEELDOEKFEVDENFTENTYLTDCCEKD 86

DB 40 DLDRG---CQTHDHCY-----NQAKKLESCKFLIDNPYNTYSYKC----- 77

QY 87 AGNAGTGDESDEVD 101
DB 78 SGNVITCSDKNNDC 92

RESULT 14

US-09-362-230-41
Sequence 41, Application US/09362230
Patent No. 6352849

GENERAL INFORMATION:

APPLICANT: Tischnfield, Jay A.

APPLICANT: Sellhammer, Jeffrey J.

TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide

TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences

TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide

TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &

STREET: 200 East Broward Boulevard

CITY: Fort Lauderdale

STATE: FL

COUNTRY: USA

ZIP: 33301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/362,230

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/888,497

FILING DATE:

APPLICATION NUMBER: US 08/097,354

FILING DATE: 26-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Manso, Peter J.

REGISTRATION NUMBER: 32,264

REFERENCE/DOCKET NUMBER: IN21044-5

TELECOMMUNICATION INFORMATION:

TELEPHONE: 305-527-2498

TELEFAX: 305-764-4996

INFORMATION FOR SEQ ID NO: 41:

SEQUENCE CHARACTERISTICS:

LENGTH: 124 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-362-230-41

Query Match 11.7%; Score 67; DB 4; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKTKCTKTEYCYIYEDTVNKECYNVVGDGEELDOEKFEVDENFTENTYLTDCCEKD 86
DB 40 DLDRG---CQTHDHCY-----NQAKKLESCKFLIDNPYNTYSYKC----- 77

QY 87 AGNAGTGDESDEVD 101
DB 78 SGNVITCSDKNNDC 92

RESULT 15
PCT-US94-07926-41

Tue May 6 16:59:48 2003

us-09-936-737a-2.rai

Page 7

```

; Sequence 41: Application PC/TUS9407926
; GENERAL INFORMATION:
; APPLICANT: Tischfield, Jay A.
; APPLICANT: Sellhammer, Jeffrey J.
; TITLE OF INVENTION: Mammalian Phospholipase A2 Nuclease
; TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences
; TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
; TITLE OF INVENTION: Sequences Having Internal Ribosome Binding Sites
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESSEE: Russell PA
; STREET: 200 East Broward Boulevard
; CITY: Fort Lauderdale
; STATE: FL
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07926
; FILING DATE: 15-JUL-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/097,354
; FILING DATE: 26-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: IN21044-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305-527-2498
; TELEFAX: 305-764-4996
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; CTT-US94-07926-41

```

	Query Match	11.7%	Score 67;	DB 5;	Length 124;
	Best Local Similarity	22.7%*	Pred. No.	5.4,	
	Matches 17; Conservative	14;	Mismatches	22;	Gaps 3,
QY	27 DLDECKKTFEKRYCYIVFEDYNNECCYYNVVDEELDGEKFVDENFTNTYTLDCEGD	86			
	: :	:	: : :	:	
Dd	40 DLRG---CQTIDHCX-----NQAKLSECFLLDNPTNVTYSKC----	77			
QY	87 AGNAAGTGSDESDD	101			
Dd	78 SGAVITCSDRNDCE	92			

Search completed: May 5, 2003, 17:51:12
Job time : 23 secs

1
:
:


```
FILE REFERENCE: 480140.434D1
CURRENT APPLICATION NUMBER: US/09/989,903
CURRENT FILING DATE: 2002-04-11
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 172
TYPE: PRF
ORGANISM: Mus musculus
US-09-989-903-16

Query Match
Best Local Similarity 35.6%; Score 75.5; DB 10; Length 172;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTAKCEYVNVDEEL---DQKRYVD--ENTENTL--TDECK-DAGNAGTG 94
DB 26 VFDDLYER---NVLNGDELKIGESASFLINKAENVLENTLKTDMAGIFAGHANSQ 81
QY 95 DE-----SDRYD 101
DB 82 EQLSLQFSNDEDD 94

RESULT 3
US-09-833-790-252
Sequence 252, Application US/09833790
Patent No. US20020068288A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Secrist, Heather
APPLICANT: Mohamath, Raodoh
APPLICANT: Indrias, Carol Y.
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.512
CURRENT APPLICATION NUMBER: US/09/833,790
CURRENT FILING DATE: 2001-04-11
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 252
LENGTH: 156
TYPE: PRF
ORGANISM: Homo sapiens
US-09-833-790-252

Query Match
Best Local Similarity 12.3%; Score 70.5; DB 10; Length 156;
Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;

QY 1 EERE-----DCWTFYANKRYTDFDKS---FKSSD-----LDECKKTCFKTEYCYIV 44
DB 30 EEOEKLKRSCLYVGNLFTTEQIYELFSKSGDIKIIIMGDKMKKTA--CGCFEVE 87
QY 45 FEDTYNKECYVNVDEELDQKRYVDENTENTL--TDECK-DAGNAGTGESDEVED 103
DB 88 YTSRADAEHAKRYINGTRLDDR--TIKIDWAGFR---EGHGYGRSGGCVREHYKOD 141

RESULT 4
US-09-815-242-11522
Sequence 11522, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
```

```
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11522
LENGTH: 1167
TYPE: PRF
ORGANISM: Helicobacter pylori
US-09-815-242-11522

Query Match
Best Local Similarity 12.2%; Score 70; DB 10; Length 1167;
Matches 24; Conservative 12; Mismatches 24; Indels 18; Gaps 4;

QY 19 DKFSKSSDLECKKTCFTEYCYIVFEDTYNKECYVNVDEELDQKRYVDENTENTY 78
DB 240 DKTEQSSNDPKENQKTLFDDA---ILLQDLAN--AYVNMPTK-----LDDRYWENF 286
QY 79 LNDCEKDGAGNAGTGDE 96
DB 287 -----AKTGNIRARTLINE 299

RESULT 5
US-09-765-272-200
Sequence 200, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
```

? TELEPHONE: (301) 309-8504
 ? TELEFAX: (301) 309-8512
 ? INFORMATION FOR SEQ ID NO: 200:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 674 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: protein
 ? SEQUENCE DESCRIPTION: SEQ ID NO: 200:
 GS-09-765-272-200

```

QY      7 WTEFANKRYVDFEFSKRSDDLECKKT-----CEKTECYIVPEDEVNKECYYNVWDS 61
        | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db     322 WIGLKNRKLEKP-----SEIIKRVATSEIQISCTSKSDISTYWF-----LRFIADPV 368

Query Match          12.1% Score 69.5; DB 10, Length 674;
Best Local Similarity 23.6%; Pred. No. 30;
Matches    25; Conservative   17; Mismatches   35; Indels   29; Gaps   -5.

QY      62 ELDOEKP-----VDENFTENLYTDCGEG-----KDAGNAAGTDE 96
        | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db     369 QEDKALYLDDDLVTFKNLDDLFAITLDQYPLAAVRDGGRAHYEGQE 414
```

```

RESULT 6
US-09-993-999-7
; Sequence 7, Application US/09993999
; Patent No. US20020110891A1
GENERAL INFORMATION:
APPLICANT: Ho, I-Cheng
APPLICANT: Arm, Jonathan P.
APPLICANT: Austen, K. Frank
APPLICANT: Glimcher, Laurie H.
TITLE OF INVENTION: Phospholipase A2 Group Preferentially
FILE REFERENCE: HUT-046
CURRENT APPLICATION NUMBER: US/09/993,999
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: 60/246,316
PRIOR FILING DATE: 2000-11-06
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 146
TYPE: PR1
ORGANISM: Mus musculus
US-09-993-999-7

```

Query Match	11.9%	Score 68	DB 10	Length 146
Best Local Similarity	25.0%	Pred. No. 7	4	
Matches	15	Conservative	9	Mismatches 18; Indels 18; Gaps
QY	27	DLDECKCKCFKREKCYIYFETVAKKCYINWVGEEELDEKREYVNDENFTENTLIDCKSGKD	86	
Db	62	DLDRCK--CQTHDKCY-----SQAKLSECKCTFLIDPNNYYSKSGSE	103	

RESULT 7
 US-09-953-280-46
 : Sequence 46, Application US/09953280
 : Publication No. US20030073072A1
 : GENERAL INFORMATION:
 : APPLICANT: Haveriga, Menzo
 : APPLICANT: Vogels, Ronald
 : APPLICANT: Boutr, Abraham
 : TITLE OF INVENTION: CHIMERIC ADENOVIRUSES
 : FILE REFERENCE: 2183-4123US
 : CURRENT APPLICATION NUMBER: US/09/953,280
 : CURRENT FILING DATE: 2001-09-14
 : PRIOR APPLICATION NUMBER: US 09/348,354
 : PRIOR FILING DATE: 1999-07-07
 : NUMBER OF SEQ ID NOS: 84

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Human Adenovirus 41 Hexon Protein
US-09-953-280-46

```

[illegible]

```

RESULT 8
US-09-801-574-57
; Sequence 57, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijing Jeremy
; APPLICANT: Page, David C.
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399,2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/1187, 518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 2789
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-801-574-57

```

[illegible]

RESULT 9
 US-09-880-748-1327
 ; Sequence 1327, Application US/09880748
 ; Publication NO. US2003005937A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Antibodies that Immunospecifically Bind BlyS
 ; FILE REFERENCE: PF523
 ; CURRENT APPLICATION NUMBER: US/09/880,748
 ; CURRENT FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 60/212,210
 ; PRIOR FILING DATE: 2000-06-15
 ; PRIOR APPLICATION NUMBER: 60/240,816
 ; PRIOR FILING DATE: 2000-10-17
 ; PRIOR APPLICATION NUMBER: 60/276,248
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/277,379

```
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1327
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1327

Query Match
Best Local Similarity 11.3%; Score 65; DB 9; Length 257;
Matches 21; Conservative 17; Mismatches 52; Indels 6; Gaps 2;

QY 8 TTYANRKYTDPEKSKSSDLD-ECKKTCFTEYCIYFEDTVNKEC-----TYNVYDGE 61
DB 53 TTYRSKWYNDYAVSVKSRITIMPDTSKNQFSIKLSTYAADTAVYVCARERPYDILTGX 112

QY 62 ELDQEKFYVDENFTENYLTDCGKAGNAGTDES 97
DB 113 TTYGMDVWGQGTIVTSSGGGSGGGSGGSOS 148

RESULT 10
US-09-974-298-114
; Sequence 114, Application US/09974238
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 3267030CD1
US-09-974-298-114

Query Match
Best Local Similarity 11.3%; Score 65; DB 9; Length 1125;
Matches 24; Conservative 24; Mismatches 39; Indels 18; Gaps 4;

QY 1 EEREDCWTYANRKYTDPEKSKSSDLD-ECKKTCFTEYCIYFEDTVNKECYNVVDG 60
DB 1032 EEDDED-----SEKEEEDKEKEEJOEKEKEPOGDEE-----EEEEEE-----VEE 1075

QY 61 EELDQEKFYVDENFTENYLTDCGKAGNAG--TGDESDVEDVD 103
DB 1076 EEEVEAENGESEKTEGLMKDDRAESQASLSQKVGESSEGVSEE 1120

RESULT 11
US-09-975-456B-5
; Sequence 5, Application US/09975456B
; Publication No. US20030073087A1
; GENERAL INFORMATION:
; APPLICANT: LAZDUNSKI, MICHEL
; APPLICANT: LAMBEAU, GERARD
; APPLICANT: VALENTIN, EMMANUEL
; TITLE OF INVENTION: NOVEL MAMMALIAN SECRETED GROUP IIF PHOSPHOLIPASE A2
; FILE REFERENCE: 1478-R-00
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: 60/239,491
```

```
; PRIOR FILING DATE: 2000-10-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln version 2.1
; SEQ ID NO 5
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-975-456B-5

Query Match
Best Local Similarity 11.3%; Score 64.5; DB 9; Length 148;
Matches 17; Conservative 5; Mismatches 20; Indels 15; Gaps 2;

QY 28 LDCEKTCFTEYCIYFEDTVNKECYNVVDGEELDQEKFYVDENFTENYLTDCG 84
DB 60 VDLELDKCCQTHDNCY-----DQAKK-----LDSCKFLDNPYTHYSYSCSG 101

RESULT 12
US-09-925-297-511
; Sequence 511, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 511
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (156)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-511

Query Match
Best Local Similarity 11.3%; Score 64.5; DB 10; Length 156;
Matches 17; Conservative 5; Mismatches 20; Indels 15; Gaps 2;

QY 28 LDCEKTCFTEYCIYFEDTVNKECYNVVDGEELDQEKFYVDENFTENYLTDCG 84
DB 68 VDLELDKCCQTHDNCY-----DQAKK-----LDSCKFLDNPYTHYSYSCSG 109

RESULT 13
US-10-102-806-695
; Sequence 695, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 695
; LENGTH: 460
```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:48:57 ; Search time 36 Seconds
(without alignments)
275.051 Million cell updates/sec

Title: US-09-936-737A-2

Sequence: 1 EEREDCWTIFYANRKYTDFDK.....GKDAGNAAGTGDESDVEDED 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Maximum DB seq Length: 200000000000

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Listing first 45 summaries

Database :

```
PIR_73:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	80.5	14.0	147	2	A42435	Ieoch antiplatelet
2	80.5	14.0	412	2	A55320	immunophilin FKBP4
3	78.5	13.7	550	2	T26562	hypothetical protein
4	78.5	13.7	907	2	T27317	hypothetical protein
5	78	13.6	682	2	B84021	two-component sensor
6	78	13.6	797	2	A36811	hypothetical protein
7	75.5	13.2	264	2	T22054	hypothetical protein
8	75.5	13.2	491	2	T16354	hypothetical protein
9	75	13.1	360	2	T34510	hypothetical protein
10	75	13.1	3844	2	T18402	hypothetical protein
11	74.5	13.0	1277	2	T32731	asparagine/aspartate
12	74	12.9	296	2	D70465	pAK interacting protein
13	74	12.9	1278	2	B70236	hypothetical protein
14	74	12.9	4466	1	S17231	dynein beta heavy
15	73	12.7	643	2	T21428	hypothetical protein
16	73	12.7	1405	1	D172PA	DNA-directed DNA p
17	73	12.7	2543	2	T31687	surface antigen - p
18	72.5	12.7	564	2	T25945	hypothetical protein
19	72.5	12.7	2533	2	T28675	alpha-5ID immobiliza
20	72.5	12.7	2533	2	T28674	alpha-5ID-immobiliz
21	72	12.6	1303	2	B84493	probable Aphbia re
22	71.5	12.5	320	2	A71611	Ser/Thr protein kin
23	71.5	12.5	594	2	F81359	para-aminobenzoate
24	71	12.4	782	2	S50719	hypothetical protein
25	70.5	12.3	146	1	PSDG	phospholipase A2 (
26	70.5	12.3	156	2	I37222	diatomic cap bindin
27	70.5	12.3	156	2	G60109	nuclear cap bindin
28	70.5	12.3	256	2	G96774	hypothetical protein
29	70.5	12.3	748	2	C64409	hypothetical protein

30	70.5	12.3	786	2	C86406	88.6k hypothetical protein
31	70	12.2	324	2	T24432	hypothetical protein
32	70	12.2	1058	2	S50295	hypothetical protein
33	70	12.2	1167	2	F71909	hypothetical protein
34	70	12.2	1255	2	B97104	hypothetical protein
35	69.5	12.1	146	1	PSBGA	phospholipase A2 (class 1)
36	69.5	12.1	452	2	T37899	hypothetical protein
37	69.5	12.1	658	2	S61160	hypothetical protein
38	69.5	12.1	636	2	D95206	hypothetical protein
39	69.5	12.1	738	2	A13625	ribonucleoside-diphosphate kinase
40	69	12.0	164	2	T33824	hypothetical protein
41	69	12.0	374	2	T33328	hypothetical protein
42	69	12.0	1137	2	T28317	ORF MSV156 hypothetical protein
43	69	12.0	1650	2	T18444	hypothetical protein
44	68.5	12.0	385	2	T26487	hypothetical protein
45	68.5	12.0	597	2	D70100	phosphoglucosaminidase

ALIGNMENTS

RESULT 1

leech antiplatelet protein precursor - Mexican leech

C;species: Haementeria officinalis (Mexican Leech)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 07-May-1999

R; Keller, P.M.; Schultz, L.D.; Condra, C.; Karczewski, J.; Connolly, T.M.

A; Title: An inhibitor of collagen-stimulated platelet activation from the salivary gland of *Phyllago*...

A;Accession: A42435

A;Residues: 1-147 <KEL>

A; Note: the authors translated the codon AGC for residu

A; Note: sequence extracted from NCBI backbone (NCBIN:91615) and modified

J. Biol. Chem. 267, 6893-6898, 1992

A; Reference number: A42434; MUID:92202246; PMID:1551897

A;Molecule type: protein

```
F:1-21/Domain: signal sequence #status predicted <SIG>
```

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Best Local Similarity 32.7%; Pred. No. 1.7;

1. **Introduction**

[illegible]

RESULT 2

Immunophilin FKBP46 - fall armyworm

C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 02-Sep-2000

R; Alnemri, E.S.; Fernandes-Alnemri, T.; Pomeroy, K.; Robertson, N.M.; Dudley, K.; D

A;Title: FKBP46, a novel Sf9 insect cell nuclear immunophilin that forms a protein-k1

A;Accession: A55320

A; molecule type: mRNA

A;Cross-references: GB:U15038; NID:g595844; PID:g595845

QY 67 KEVDENFTENYLTDCGKAGNAGTGDSEDEVD 102
 Db 428 KDEKEGEDEGDDEDEDE--DEDEGDEGDE 461

RESULT 7

T22054
 hypothetical protein F40G12.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
 C:Accession: T22054
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19507
 A:Accession: T22054
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-264 <TAL>
 A:Cross-references: EMBL:Z77661; PIDN:CA801183.1; GSPDB:GN00023; CESP:F40G12.5
 A:Experimental source: clone F40G12
 C:Genetics:
 A:Gene: CESP:F40G12.5
 A:Map position: 5
 A:Introns: 57/3; 158/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein F40G12.4

Query Match 13.2%; Score 75.5; DB 2; Length 264;
 Best Local Similarity 27.7%; Pred. No. 6.8;
 Matches 31; Conservative 19; Mismatches 33; Indels 29; Gaps 8;

QY 1 EEREDCWTFFYANKRYTDPDKSFK--SPDLDECKKTCFTEKYIYFE-----DTYNKE 52
 Db 29 ESEEDC-----FEKVFLLIISCKHECSK-----DYDFLRNLIQRREALTSGKE 72

QY 53 CYYNVDEBELDOEKF-VVDENFTE--NYLTDCGKAGNAGTGDSEDEVD 101
 Db 73 CFLEIVK-EECPKEKFLIEENYSQVLLTE-KPKNGACTAPYFQLEIE 122

RESULT 8

T16354
 hypothetical protein F42G9.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16354
 R:Taich, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of C. elegans cosmid F42G9.
 A:Reference number: Z18498
 A:Accession: T16354
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-491 <TAL>
 A:Cross-references: EMBL:U00051; NID:G1216305; PID:G1216310; PIDN:AAA91358.1; CESP:F42G9
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F42G9.1
 A:Introns: 37/3; 254/2; 364/2; 406/2

Query Match 13.2%; Score 75.5; DB 2; Length 491;
 Best Local Similarity 24.5%; Pred. No. 16;
 Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

QY 10 YANRKYTF--DKSFKSSDLDECKTCFTEKYIYFEEDYVNEKYCYNVVDEBELDOEK 67
 Db 67 FTSKLPFLFKKFWEDADVAECLQKAF-----VDFDFRAE-----ESMKELR 112

QY 68 FVVDENFTENYLTDCGKAGNAGTGDSEDEVD 101
 Db 113 DIGDEG-----KKPKAGGADSDSEADRID 137

RESULT 9

T34510
 hypothetical protein ZK1290.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T34510
 R:Taich, A.
 submitted to the EMBL Data Library, July 1995
 A:Description: The sequence of C. elegans cosmid ZK1290.
 A:Reference number: Z21535
 A:Accession: T34510
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-360 <TAL>
 A:Cross-references: EMBL:U21308; PIDN:AA93316.1; GSPDB:GN00020; CESP:ZK1290.10
 A:Experimental source: strain Bristol N2; clone ZK1290
 C:Genetics:
 A:Gene: CESP:ZK1290.10
 A:Map position: 2
 A:Introns: 23/1; 48/1; 158/3; 216/1; 264/1; 312/1; 331/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1290.10

Query Match 13.1%; Score 75; DB 2; Length 360;
 Best Local Similarity 33.3%; Pred. No. 13;
 Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;

QY 46 EDYVNEKYCYNVVDEBELDOEKFYV-----DENFTENYLTDCGKAGNAGTGDSEDEVD 101
 Db 224 EDYVNEKYCYNVVDEBELDOEKFYV-----DENFTENYLTDCGKAGNAGTGDSEDEVD 101

QY 95 DESDEVED 103
 Db 281 DDEDEDD 289

RESULT 10

T18402
 aspartate/aspartate rich protein - malaria parasite (Plasmodium falciparum) (fragment)
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18402
 R:Barale, J.C.; Candellie, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley, R.
 Infect. Immun. 65, 3003-3010, 1997
 A>Title: Plasmodium falciparum AAP1, a giant protein containing repeated motifs rich
 A:Reference number: Z18923; MIMD:97378065; PMID:9234746
 A:Accession: T18402
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3844 <BAR>
 A:Cross-references: EMBL:Y08926; NID:el154302; PID:el1435; PIDN:CAA70130.1
 C:Genetics:
 A:Gene: aarp1

Query Match 13.1%; Score 75; DB 2; Length 3844;
 Best Local Similarity 23.9%; Pred. No. 1.5e+02;
 Matches 21; Conservative 23; Mismatches 34; Indels 10; Gaps 3;

QY 1 EEREDCWTFFYANKRYTDPDKSFK--SPDLDECKKTCFTEKYIYFEEDYVNEKYCYNVVDEBELDOEK 67
 Db 1726 EDSEECQIKSYKMSKCNKNNIIFDSISVLRKNNIKRLKNYMKCNKNCIYIDNNRK 1785

QY 52 EGYNVVDEBELDOEKFWVDENFTENYLT 79
 Db 1786 K-KNNKKNVNDKEKRYVLLKIFVHNFI 1812

RESULT 11

T32731
 PAR interacting protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32731
 R:Comte, P.A.; Ossipow, V.; Schibler, U.
 submitted to the EMBL Data Library, January 1997

A:Description: Isolation of PIP, a 160 kDa nuclear protein that interacts with the act
A:Reference number: Z21213
A:Accession: T32731
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1277 <CON>
A:Cross-references: EMBL:U83590; NID:g2253210; PID:g2253211; PIDN:AA62878.1

Query Match 13.0%; Score 74.5; DB 2; Length 1277;
Best Local Similarity 30.4%; Pred. No. 53;
Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;

QY 1 EEREDCTFYANKRYTDFD-KSKFKSSDLDECKKTCFTEYCYIVFEDTVNKECYNVVD 59
DB 630 EDEED-----NNVVFDTDEKOLKHGEDADS-----EDSKSESDVSDSD 670
QY 60 GEELDOE--KVVVDENFENVLTDCEKGDAGNAGTGDSEDE 99
DB 671 GESEEDRDKVDYDPRQOLM--EVLQAGNALGGESEEE 709

RESULT 12
D70465
hypothetical protein aq_1923 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
A:Accession: D70465
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-296 <AGF>
A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07719.1; PID:g2984187; GB:AE00065
A:Experimental source: strain VFS
C:Genetics:
A:Gene: aq_1923
C:Superfamily: Aquifex aeolicus hypothetical protein aq_1923

Query Match 12.9%; Score 74; DB 2; Length 296;
Best Local Similarity 24.2%; Pred. No. 14;
Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

QY 9 FYANRKYTD-----FDKSPFKSSDLDECKKTCFTEYCYIVE---DVKVKECY 54
DB 18 FYATDKYMDNKAQVEYFLDKTLRKSG-----KGSYKYVDYKPIGGELIHKDY 67
QY 55 YNVVDGEELDOEKFVVDENFENVLTDCEK 85
DB 68 YRDRNGEEFKIEELIIEK-----LSFECK 92

RESULT 13
B70236
hypothetical protein BBH09 - Lyme disease spirochete plasmid H/Jp28-3
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
A:Accession: B70236
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1278 <RLF>
A:Cross-references: GB:AE000784; NID:g2690041; PIDN:AAC66000.1; PID:g2690056; TIGR:BBH09

A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 12.9%; Score 74; DB 2; Length 1278;
Best Local Similarity 25.0%; Pred. No. 59;
Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

QY 10 YANRRYTFDKSPFKSSDLDECKKTCFTEYCYIVFEDTVNKECYNV-----VDGEELD 64
DB 181 YEKRYIED--FSKIKKEE-----YKEGEWFLFSYLIRKRYLKTSLIISVEKEQIS 233
QY 65 QEFVVDENFTE 76
DB 234 KEKEIIORTLRE 245

RESULT 14
S17231
dynein beta heavy chain, ciliary - sea urchin (Anthocidaris crassispina)
N:Contains: dynein ATPase (EC 3.6.4.2)
C:Species: Anthocidaris crassispina
C:Date: 30-Sep-1991 #sequence_revision 02-May-1994 #text_change 19-Apr-2002
A:Accession: S17231; PS0415
R:Ogawa, K.
Nature 352, 643-645, 1991
A:Title: Four ATP-binding sites in the midregion of the beta heavy chain of dynein.
A:Reference number: S17231; MUID:91326104; PMID:1830928
A:Accession: S17231
A:Molecule type: mRNA
A:Residues: 1-4466 <OGA>
A:Cross-references: GB:D01021; NID:g217202; PIDN:BA00827.1; PID:g217203
R:Ogawa, K.
Proc. Jpn. Acad. B Phys. Biol. Sci. 67, 27-31, 1991
A:Title: ATP-binding site in dynein beta-heavy chain: identification by molecular cloning.
A:Reference number: PS0415
A:Accession: PS0415
A:Molecule type: mRNA
A:Residues: 764-1001, 'APQ', 1005-2036, 'VPSSVET' <OG2>
C:Superfamily: dynein heavy chain, ciliary
C:Keywords: ATP; heterotetramer; hydrolase; microtubule binding; nucleotide binding;
F:154-161/Region: nucleotide-binding motif A (P-loop)
F:1852-1859/Region: nucleotide-binding motif A (P-loop)
F:2133-2140/Region: nucleotide-binding motif A (P-loop)
F:2460-2467/Region: nucleotide-binding motif A (P-loop)
F:2805-2812/Region: nucleotide-binding motif A (P-loop)
F:160/Binding site: ATP (Lys) #status predicted
F:1858/Binding site: ATP (Lys) #status predicted
F:2133/Binding site: ATP (Lys) #status predicted
F:2466/Binding site: ATP (Lys) #status predicted
F:2811/Binding site: ATP (Lys) #status predicted

Query Match 12.9%; Score 74; DB 1; Length 4466;
Best Local Similarity 25.9%; Pred. No. 21e+02;
Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;

QY 1 EEREDCTFYANKRY-----TDFKSPFKSSDLDE-----CK-----KTC 35
DB 460 EEEQECATKFTETPRYDGLPTCOEFLDEYEEFEKFFDLDRGLGILCOGFPDCCGLEMA 519
QY 36 FTEYCY-----YFETVKNKECYNV-----DGEELDOEKFVVDEN 73
DB 520 FKMLDCYGLDPRVIRNDF-----ECKYPTVIMLYDQELDQSKREIYDH 564

RESULT 15
T21428
hypothetical protein F26F2.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A:Accession: T21428
R:Ainscough, R.
submitted to the EMBL Data Library, November 1996

Tue May 6 16:59:49 2003

us-09-936-737a-2.rpr

Page 5

A; Reference number: Z19420

A; Accession: T21428
A; Status: preliminary

A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA

A;Molecule type: DNA
A:Peptides: 1-643

A;Residues: 1-643 <WIL>
A;Cross-references: FMB

A/Cross-references: EMBL:Z81514; PIDN:CAB04192.1; GSPDB:GN00023; CESP:F26F2.6
A/Experimental source: clone F26F2

A; Experimental source: clone F26F2 C: Genetics.

C; Genetics: A: Gene: CFS

A:Gene: CESP:F26F2.6
A:Map position: 5

A:Map position: 5
A:Introns: 18/3; 116/3; 142/1; 265/3; 306/3; 373/3; 403/2; 425/2; 479/3; 544/3; 573/2; 6

Quarry Match 12 78: Score 73: DB 3: Tonath 643:

Query Match
Best: Local

Best Local Similarity 26.6%; Pred. No. 37;
Matches 29: Conservative 14: Mismatches 32: Indels

Matches	29;	Conservative	14;	Mismatches	32;	Indels	34;	Gaps	6;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

QY 7 WIFY-----ANRKYT---DFDKSFKS-----SDLDECKKTCFTEYCYIVFED 47

Db 155 WSEFRPYVDGCNANCSVTMVQDYAKPGPETVKPRDAQISNWDECMHMCYEEECQIAYLD 214

QY 48 TVNKEC-YYNVGDGEELDÖ-----KFVDEN---FTENYLT D 81

Db 215 SDAKCVWYSSDDGLTFMNSSADSGKRLVIKMKVDETTCLFTTPQLLD 263

Search completed: May 5, 2003, 17:53:28
Job time : 40 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2003, 17:20:27 ; Search time 21 seconds

(Without alignments)
203.432 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EERDCWTFYANKRYTDFDK.....GKDAAGNAGTGDSEDEVDD 103

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	85.5	14.9	147	1 LAP_HABOF	Q01747 haementeria
2	80.5	14.0	412	1 FEB4_SPOFR	Q26486 spodiopoda
3	78	13.6	797	1 VG48_HSVSA	Q01033 herpesvirus
4	75.5	13.2	419	1 ICEC_MOUSE	Q08736 mus musculus
5	75.5	13.2	491	1 P2C1_CAEEL	P49595 caenorhabditis
6	75	13.1	360	1 YOP4_CAEEL	Q09337 caenorhabditis
7	74	12.9	296	1 Y023_AQDAE	Q67757 aquilex aeo
8	73	12.9	4466	1 DHQC_ANTCR	P39057 anthracis
9	73	12.7	1405	1 DPOA_SCHPO	P38040 schistosoma
10	71	12.4	782	1 YNV2_YEAST	P40157 saccharomyces
11	70.5	12.3	78	1 MOH1_CANGP	P40157 cancer pagu
12	70.5	12.3	146	1 PA21_CANGP	P06596 canis familiaris
13	70.5	12.3	156	1 CB20_HUMAN	P52288 homo sapiens
14	70.5	12.3	748	1 Y875_METVA	Q58285 methanococcus
15	70	12.2	1058	1 S185_YEAST	P40856 saccharomyces
16	69.5	12.1	146	1 PA21_PIG	P00592 sus scrofa
17	68.5	12.0	610	1 CALG_HUMAN	Q14967 homo sapiens
18	68	11.9	146	1 PA21_MOUSE	Q92092 mus musculus
19	68	11.9	240	1 NEF_HV2D2	P15829 human immunodeficiency
20	68	11.9	4466	1 DYHC_TRIGR	P23098 tritipeustus
21	67.5	11.8	277	1 PS12_ARATH	Q33712 arabidopsis
22	67.5	11.8	321	1 D7_AEDAE	P18153 aedes aegypti
23	67.5	11.8	602	1 SPT8_YEAST	P38915 saccharomyces
24	67	11.7	146	1 PA21_RAT	P04055 rat mus norvegicus
25	67	11.7	327	1 RRP_PIRYV	Q01769 piry virus
26	67	11.7	650	1 CC45_YEAST	Q08033 saccharomyces
27	67	11.7	2410	1 MOH1_SCHPO	Q08033 schistosoma
28	66.5	11.6	78	1 MOH2_CANGP	P40157 cancer pagu
29	66.5	11.6	125	1 FABL_ANOPU	Q90238 anolis pulchellus
30	66.5	11.6	132	1 PA21_HORSE	P00594 equus caballus
31	66.5	11.6	405	1 LMX1_BRANE	Q90476 brachydanio rerio
32	66.5	11.6	591	1 CALX_RAT	P35565 rat mus norvegicus
33	66.5	11.6	733	1 YV08_YEAST	P46995 saccharomyces

ALIGNMENTS

RESULT 1	LAP_HABOF	STANDARD	PRT	147 AA
AC	Q01747			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DE	Anti-platelet protein precursor.			
GN	LAP.			
OS	Haementeria officinalis (Mexican leech).			
OC	Eukaryota; Metazoa; Annelida; Clitellata; Hirudini; Hirudinea;			
OC	Rhynchobdellida; Glossiphoniidae; Haementeria.			
OX	NCBI_TaxID=6410;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 60-91 AND 123-139.			
RC	TISSUE-Salivary gland.			
RX	MEDLINE=92202247; PubMed=1551898;			
RA	Keller P.M., Schultz L.D., Condra C., Karczewski J., Connolly T.M.;			
RT	"An inhibitor of collagen-stimulated platelet activation from the			
RT	salivary glands of the Haementeria officinalis leech. II. Cloning of			
RT	the cDNA and expression."			
RT	J. Biol. Chem. 267:6899-6904 (1992)			
CC	- FUNCTION: AN INHIBITOR OF COLLAGEN-STIMULATED PLATELET			
CC	AGGREGATION, DENSE GRANULE RELEASE AND SEROTONIN RELEASE.			
CC	- SUBCELLULAR LOCATION: Secreted.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
DR	EMBL; M81489; AAA29194.1; -			
KW	SIGNAL.			
FT	SIGNAL 1 21			POTENTIAL.
FT	CHAIN 22 147			ANTI-PLATELET PROTEIN.
FT	SEQUENCE 147 AA; 15908 MW; 75A55113744EA2E CRC64;			
QY	Query Match	14.98;	Score 85.5;	DB 1; Length 147;
QY	Best Local Similarity	33.98;	Pred. No. 0.2;	
QY	Matches 19; Conservative 10; Mismatches 24; Indels 3; Gaps 1;			
Db	65 EDCWTFYANKRYTDFDKSFKSSDLECKTCFETK---CYIYEDTVAKKCYNN 56			
Db	EDCWSKRPGLPDLNLLTKTEFTSVDCRCKRCSAVEPSCYIIQINTNECYRN 120			
RESULT 2	FKB4_SPOFR	STANDARD	PRT	412 AA
AC	Q26486			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			

```

DE 46 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
DE isomerase) (Ppiase) (EC 5.2.1.8).
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP MEDLINE=95074110; PubMed=7527037;
RA Alnemri E.S., Fernandes-Alnemri T., Pomeroy K., Robertson N.M.,
RA Dudley K., Dubois G.C., Litwack G.;
RT "FKBP46, a novel sf9 insect cell nuclear immunophilin that forms a
RT protein-kinase complex."
RL J. Biol. Chem. 269:30828-30834(1994).
CC -1- FUNCTION: BINDS TO, AND IS INHIBITED BY FK506 AND RAPAMYCIN. BINDS
CC DOUBLE-STRANDED DNA IN VITRO. PPIASES ACCELERATE PROTEIN FOLDING.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED BY A NUCLEAR KINASE IN THE PRESENCE OF MG2+
CC AND ATP.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U15038; AAA58962.1; -
DR HSSP: P20071; 1FKJ.
DR InterPro: IPR001179; FKBP_PPIase.
DR Pfam: PF00254; FKBP; 1.
DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
KM Isomerase; Rotamase; Nuclear protein; DNA-binding; Phosphorylation.
FT DOMAIN 91 112 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 120 145 LYS-RICH (BASIC).
FT DOMAIN 152 216 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 219 302 LYS-RICH (BASIC).
FT DOMAIN 324 412 PPIASE, FKBP-TYPE.
SQ SEQUENCE 412 AA; 45810 MW; F2A69159AEF4FE22 CRC64;
Query Match 14.0%; Score 80.5; DB 1; Length 412;
Best local Similarity 30.6%; Pred. No. 1.7;
Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;
QY 11 ANKRYDFOKSEFKS-----SDLECKKTCFETCYIVFEDTVNKECYVVDGEELDQ 65
DB 133 ANKRAPDKKAGKNSAPAEASDSDDD-----EDQLK-----FLDGEDIDT 174
DY 66 EKFEVDNFTENYLTDCGKDGAGNAGTGDESDVEVD 103
DB 175 DE--NDSFKRM--TSAEGDDSDDEDDDEDDDD 208

```

```

RN [1]
RP MEDLINE=9233688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Bisinger B.,
RA Newman C., Wilmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome."
RL J. Virol. 66:5047-5058(1992).
CC -1- SIMILARITY: TO EBV BRRE2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: X64346; CAA45671.1; -
DR PIR: A36811; A36811.
KM Hypothetical protein.
SQ SEQUENCE 797 AA; 88667 MW; 9E294234AD850E23 CRC64;
Query Match 13.6%; Score 78; DB 1; Length 797;
Best local Similarity 32.3%; Pred. No. 5.7;
Matches 31; Conservative 12; Mismatches 33; Indels 20; Gaps 5;
QY 11 ANKRYDFOKSEFKS-----SDLECKKTCFETCYIVFEDTVNKECYVVDGEELDQ 66
DB 362 ANEK--EKRIIISDDDDDRDKDEYLE-----NEE--YNRDEEDEDDEDE 427
DY 67 KFEVDNFTENYLTDCGKDGAGNAGTGDESDVEVD 102
DB 428 KDEKEGEDEGDDGDEDEGE--DEGDEDEGE 461

```

```

RESULT 4
ID ICEC_MOUSE STANDARD; PRT; 419 AA.
AC 008736;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Caspase-12 precursor (EC 3.4.22.-).
GN CASP12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP MEDLINE=97190206; PubMed=9038361;
RA van de Graen M., Vandenabeele P., Declercq W., van den Brande I.,
RA Fiers W.;
RT "Characterization of seven murine caspase family members."
RL FEBS Lett. 403:61-69(1997).
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
CC RESPONSIBLE FOR APOPTOSIS EXECUTION (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS (POTENTIAL).
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN SKELETAL MUSCLE AND LUNG.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----

```

DR EMBL: Y13090; CAA73532.1; -
DR HSSP: P29466; 1ICE.
DR MEROPS: C14.013; -
DR MGD: MG1:1312922; Casp12.
DR InterPro: IPR001315; CARD.
DR InterPro: IPR002398; ICE.
DR InterPro: IPR002138; ICE_p10.
DR InterPro: IPR001309; ICE_p20.
DR Pfam: PF00655; ICE_p10; 1.
DR Pfam: PF00656; ICE_p20; 1.
DR PRINTS: PR00376; ILBCEZYME.
DR SMART: SM00114; CARD; 1.
DR SMART: SM00115; CASc; 1.
DR PROSITE: PS0209; CARD; 1.
DR PROSITE: PS0122; CASPASE_CYS; 1.
DR PROSITE: PS0121; CASPASE_HIS; 1.
DR PROSITE: PS0207; CASPASE_P10; 1.
DR PROSITE: PS0208; CASPASE_P20; 1.
KW Hydrolyase; Thiol protease; Apoptosis; Zymogen.
FT PROPEP 1 ? ?
FT CHAIN ? ? CASPASE-12 SUBUNIT 1.
FT PROPEP ? ? POTENTIAL.
FT CHAIN ? ? CASPASE-12 SUBUNIT 2.
FT DOMAIN ? 419
FT ACT_SITE 1 92
FT ACT_SITE 250 250 CARD.
FT ACT_SITE 298 298 BY SIMILARITY.
SQ SEQUENCE 419 AA; 47853 MW; B94B0FED161CB40 CRC64;

Query Match 13.2%; Score 75.5; DB 1; Length 419;
Best Local Similarity 35.6%; Pred. No. 5;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDYVKECYNYVDGEEL---DQEKFYV--TDCEK--DAGNAGNG 94
DB 26 VFDDLYEK---NVNGDELKIGESASFLKAKENLVENLEKTDMAKIRAGHANSQ 81
QY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 5
P2C1_CAEEL STANDARD; PRT; 491 AA.
AC P49595;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable protein phosphatase 2C P4269.1 (EC 3.1.3.16) (PP2C).
GN P4269.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Tatch A.; Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
CC phosphate.
CC -1- CORACOR: BINDS 2 MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PP2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: U00051; AAA91358.1; -

DR HSSP: P35813; 1A60.
DR WormPep: F4269.1; CE07231.
DR InterPro: IPR001932; PP2C-1ike.
DR InterPro: IPR000222; PP2C.
DR Pfam: PF00481; PP2C; 1.
DR SMART: SM00331; PP2C_SIG; 1.
DR SMART: SM00332; PP2C; 1.
DR PROSITE: PS01032; PP2C; 1.
KW Hypothetical protein; Hydrolyase; Magnesium; Manganese.
FT METAL 37 37 MANGANESE 1 (BY SIMILARITY).
FT METAL 38 38 MANGANESE 1 (BY SIMILARITY).
FT METAL 57 57 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 428 428 MANGANESE 2 (BY SIMILARITY).
FT METAL 477 477 MANGANESE 2 (BY SIMILARITY).
SQ SEQUENCE 491 AA; 53141 MW; F110D12E4395306 CRC64;

Query Match 13.2%; Score 75.5; DB 1; Length 491;
Best Local Similarity 24.5%; Pred. No. 5.9;
Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

QY 10 YANKRYTF--DKSFKSSDDECKTKCFKECYIVFEDYVKECYNYVDGEELDOEK 67
DB 67 FTSAPLPFLERKRFWEADVAECLQKAF-----VFDDFIRAE-----ESMKELK 112
QY 68 FVVDENFENYITDCEGKADGNAAGTGDESDVD 101
DB 113 DIGDEG-----KKKAGGEADSEADRID 137

RESULT 6
YOFA_CAEEL STANDARD; PRT; 360 AA.
ID YOFA_CAEEL
AC Q09337;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 39.7 kDa protein ZK1290.10 in chromosome II precursor.
GN ZK1290.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Tatch A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: U21308; AAB93316.1; -
DR WormPep: ZK1290.10; CE02080.
DR InterPro: IPR003582; SHKT.
DR SMART: SM00254; SHKT; 1.
KW Hypothetical protein; Signal.
FT SIGNAL 1 21
FT CHAIN 22 360
SQ SEQUENCE 360 AA; 39669 MW; 18751B6C062DAF7 CRC64;
Query Match 13.1%; Score 75; DB 1; Length 360;
Best Local Similarity 33.3%; Pred. No. 4.8;
Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;
QY 46 EDYVKECYNYVDGEELDEKFFV-----DENFTENYITDCEGKADGNAAGTG 94
DB 224 EDYVGE---NVFDEKLFKFKIYIRNDGSTTENTTEQSTTEKTRKGIYEVSVLEGG 280

```

OY 95 DESDEVED 103
DB 281 DDEEDDD 289

RESULT 7
YU23_AQUAE STANDARD; PRT: 296 AA.
AC 067752;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein AQ_1923 precursor.
GN AQ_1923.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=96196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Augay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aerolius".
RL Nature 392:353-358(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE000763; AAC07719.1;
CC KW Hypothetical protein; Signal; Complete proteome.
CC FT SIGNAL 1 20 POTENTIAL.
CC FT CHAIN 1 296
CC SEQUENCE 296 AA; 34566 MW; 9C88F4456A9449C CRC64;
SQ

Query Match 12.9%; Score 74; DB 1; Length 296;
Best Local Similarity 24.2%; Pred. No. 4.8;
Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

OY 9 FRANKYTD-----FDKSKSSDDECKTKCTKCTCYVFE-----DIYNKCY 54
DB 18 YFAYDKYMNKAKEQVEFLDKTLTKSG-----KGSYKYVDYKPIGGIITIKDY 67
OY 55 YNVVDEELDQKFEVDENFTENYLTDEGK 85
DB 68 YDRNGEERKIEIIEIK-----LSETEGK 92

RESULT 8
DYHC_ANTCR STANDARD; PRT: 4466 AA.
AC P39057;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
OS Dyein beta chain, ciliary.
DE Anticardiac crassispina (sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Echinoidae; Echinacea; Echinoida; Echinometrididae;
OC Anticardiac.
OX NCBI_TaxID=7629;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91326104; PubMed=1830928;
RA Ogawa K.;

```

```

RT "Four ATP-binding sites in the midregion of the beta heavy chain of
RT dynein".
CC Nature 352:643-645(1991).
CC CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC CC DYNEIN HAS ATPASE ACTIVITY.
CC CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
CC CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
CC CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: D01021; BAA00827.1;
CC DR PIR: S17231; S17231.
CC DR InterPro: IPR004273; Dynein_heavy.
CC KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
CC KW Coiled coil.
CC FT DOMAIN 482 502 COILED COIL (POTENTIAL).
CC FT DOMAIN 627 643 COILED COIL (POTENTIAL).
CC FT DOMAIN 734 805 COILED COIL (POTENTIAL).
CC FT DOMAIN 1036 1056 COILED COIL (POTENTIAL).
CC FT DOMAIN 1306 1337 COILED COIL (POTENTIAL).
CC FT DOMAIN 1443 1468 COILED COIL (POTENTIAL).
CC FT DOMAIN 1950 1978 MICROTUBULE-BINDING (POTENTIAL).
CC FT DOMAIN 3033 3134 COILED COIL (POTENTIAL).
CC FT DOMAIN 3263 3325 COILED COIL (POTENTIAL).
CC FT DOMAIN 3573 3642 COILED COIL (POTENTIAL).
CC FT NP_BIND 154 161 ATP (POTENTIAL).
CC FT NP_BIND 1852 1859 ATP (POTENTIAL).
CC FT NP_BIND 2133 2140 ATP (POTENTIAL).
CC FT NP_BIND 2460 2467 ATP (POTENTIAL).
CC FT NP_BIND 2805 2812 ATP (POTENTIAL).
CC SEQUENCE 4466 AA; 511772 MW; C465CC5C6C6D325D CRC64;
SQ

Query Match 12.9%; Score 74; DB 1; Length 4466;
Best Local Similarity 25.9%; Pred. No. 81;
Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;

OY 1 EEREDCTWFANKRY-----TDFKSKSSDDE-----CK-----KTC 35
DB 460 EEFQECNAKVFTEKPYGSLDPTQGFLEDEFEKVFYDLRLGSLICGFDCCGLEAA 519
OY 36 KTEYCY-----YFEDTVNKECYNNV-----DGEELDEKRVVDEN 73
DB 520 FKMLDYGPLLDREVIRNDE---ECKYPIVIMDYDELDSKREIYDEH 564

RESULT 9
DPOA_SCHPO STANDARD; PRT: 1405 AA.
AC P28040;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
OS DNA polymerase alpha catalytic subunit (EC 2.7.7.7) (DNA polymerase
DE 1).
GN POL1 OR SM17 OR SPAC3H5.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91238692; PubMed=2034212;
RA Damaguer V., Tiliat J., de Recondo A.-M., Baldacci G.;

```

```

RT  "The POLI gene from the fission yeast, Schizosaccharomyces pombe,
RT  shows conserved amino acid blocks specific for eukaryotic DNA
RT  polymerases alpha."
RL  Mol. Gen. Genet. 226:182-189(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=93140876; PubMed=8423854;
RA  Jagmohan S., Klar A.J.S.;
RT  "DNA polymerase-alpha is essential for mating-type switching in
RT  fission yeast."
RL  Nature 361:271-273(1993).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN=972;
RX  MEDLINE=21848401; PubMed=11859360;
RA  Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA  Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA  Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA  Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Fraser A.,
RA  Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Hodgson G.,
RA  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA  Mooney P., Moulé S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA  Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA  Skelton J., Simmonds M., Squares R., Stevens K.,
RA  Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA  Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA  Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA  Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA  Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA  Beger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA  Goffeau A., Cadieu E., Dreano S., Gloux S., Leleux V., Mottier S.,
RA  Calbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA  Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Rhode G.,
RA  Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA  Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA  Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT  "The genome sequence of Schizosaccharomyces pombe.";
RL  Nature 413:871-880(2002).
CC  -1- FUNCTION: POLYMERASE ALPHA IN A COMPLEX WITH DNA PRIMASE IS A
CC  REPLICATIVE POLYMERASE.
CC  -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC  + (DNA)(N).
CC  -1- SUBCELLULAR LOCATION: Nuclear
CC  -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC  ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC  DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC  -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: X58299; CAA4132.1; -
DR  EMBL: X69673; CAB57881.1; -
DR  EMBL: Z99296; CAB16598.1; -
DR  PIR: S15993; DJZPA.
DR  InterPro: IPR002064; DNA_pol_B.
DR  InterPro: IPR004578; POL2.
DR  Pfam: PF00136; DNA_pol_B; 1.
DR  Pfam: PF03104; DNA_pol_B_exo; 1.
DR  PRINTS: PR00106; DNAPOLB.
DR  SMART: SM00486; POLBc; 1.
DR  TIGRFAMs: TIGR00592; pol2; 1.
DR  PROSITE: PS00116; DNA_POLYMERASE_B; 1.
KW  Transferase; DNA-directed DNA polymerase; DNA replication;
KW  DNA-binding; Nuclear protein.

```

```

SQ  SEQUENCE 1405 AA; 159348 MW; 06082A3BAD6347C2 CRC64;
Query Match
Best Local Similarity 32.3%; Score 73; DB 1; Length 1405;
Matches 21; Conservative 10; Mismatches 18; Indels 16; Gaps 3;
QY 45 FEDYVNRKCYNNVNGE-----ELDQEKVVDN---FENYITDCEGDAGNAAGTG 94
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 31 YESKNEELYNVSEERYKRVQRDLDDDFVDDNGAGYVDN-----GYDEWDQSHYS 84
QY 95 DESDE 99
    | | |
DB 85 DEDDE 89
RESULT 10
YVW2_YEAST STANDARD; PRT; 782 AA.
ID YVW2_YEAST
AC PA0157;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 88.8 kDa protein in PEX17-MER1 intergenic region.
GN YN212W OR N1327.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RX SMPAIN-S288C; FY1679;
RX MEDLINE=95282517; PubMed=7762305;
RA Coster F., van Dyck L., Joniaux J.-L., Purnelle B., Goffeau A.;
RT "The sequence of a 13.5 kb DNA segment from the left arm of yeast
RT chromosome XIV reveals MER1, RAP1, a new putative member of the DNA
RT replication complex and a new putative serine/threonine phosphatase
RT gene.";
RL Yeast 11:85-91(1995).
CC -1- SIMILARITY: TO C.CARDUNCULUS PROTEIN CYPD4 AND S.POMBE
CC SPEC1685.14C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X78898; CAA55496.1; -
DR EMBL: Z71488; CAA96114.1; -
DR SGD: S0005156; YNL212W.
KW Hypothetical protein.
SQ SEQUENCE 782 AA; 88845 MW; CCBBI4FACE32BF97C CRC64;
Query Match
Best Local Similarity 26.5%; Score 71; DB 1; Length 782;
Matches 22; Conservative 16; Mismatches 39; Indels 6; Gaps 2;
QY 24 KSSLDDECKKTCFTECYIVFEDTVNKECYNNVNGE---LDQEKVVDNTEFVYLT 80
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 315 KFKDLDCIKCIQRFAMVNCIMW---TLNKEITVDPEREKVDIIDSSVPLEKQDDLIHI 371
QY 81 DCEGKDGNAAGTDESDVDDED 103
    | : : : : | | | |
DB 372 DDRSNEERDKESSESENDESD 394
RESULT 11
MOH1_CAMPG STANDARD; PRT; 78 AA.
ID MOH1_CAMPG
AC P81034;
DT 15-JUL-1998 (Rel. 36, Created)

```

DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Mandibular organ-inhibiting hormone-1 (MOIH-1).
 OS Cancer pagurus (Crab crab).
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
 CC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
 CC Brachyura; Eubrachyura; Cancroidea; Cancridae; Cancer.
 OX NCBI_TaxID=6755;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=stomach gland;
 RX MEDLINE=96278655; Pubmed=8662685;
 RA Malinwight G., Webster S.G., Wilkinson M.C., Chung J.S., Rees H.H.;
 RT "Structure and significance of mandibular organ-inhibiting hormone in
 the crab, Cancer pagurus. Involvement in multihormonal regulation of
 growth and reproduction.";
 RL J. Biol. Chem. 271:12749-12754(1996).
 CC -1- FUNCTION: REPRESENTS THE SYNTHESIS OF METHYL FARNESATE, THE
 PRECURSOR OF INSECT JUVENILE HORMONE III IN THE MANDIBULAR ORGAN.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE MEDULLA TERMINALIS X-ORGAN IN
 THE EYES/TALKS AND TRANSPORTED TO THE SINUS GLAND WHERE IT IS
 STORED AND RELEASED.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD CHH/MIH/GIH/VIH FAMILY OF
 HORMONES.
 CC InterPro: IPR001166; CHH_MIH_GIH.
 DR Pfam: Pf01147; Crust_neurohorm; 1.
 DR PRINTS: PR00550; HYPRGLYCEMIC.
 DR PROSITE: PS01250; CHH_MIH_GIH; 1.
 KW Neuropeptide; Hormone
 FT DISULFID 7 44 BY SIMILARITY.
 FT DISULFID 24 40 BY SIMILARITY.
 FT DISULFID 27 53 BY SIMILARITY.
 SQ SEQUENCE 78 AA; 9241 MW; 8DB338A39058A62D CRC64;

Query Match 12.3%; Score 70.5; DB 1; Length 78;
 Best Local Similarity 33.9%; Pred. No. 2.5;
 Matches 20; Conservative 5; Mismatches 23; Indels 11; Gaps 3;

OY 5 DCMTFYANRKY-----TDFDKSPKSSDLDCKKTCF-KTEY--CYVFEDYVYKE 52
 DB 6 DCQNFICRMYEKVEVDWICKNCANIFRKDGLNLCNSCYNTEFLMCIDATENTRKE 64

RESULT 12
 ID PA21_CANFA STANDARD; PRT; 146 AA.
 AC P06596;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-
 acylhydrolase) (Group IB phospholipase A2).
 GN PLA2G1B.
 OS Canis familiaris (Dog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86223862; Pubmed=3754861;
 RA Ohara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shln M.,
 RT "Dog and rat pancreatic phospholipases A2: complete amino acid
 sequences deduced from complementary DNAs.";
 RL J. Biochem. 99:733-739(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90306027; Pubmed=2142076;
 RA Kefelef B., Laforge K.S., Vasiloudes P., Puigserver A., Scheele G.A.;
 RT "Isolation and sequence of the canine pancreatic phospholipase A2
 gene.";
 RL Eur. J. Biochem. 190:299-304(1990).
 RN [3]

RP SEQUENCE FROM N.A.
 RX MEDLINE=87175472; Pubmed=3562437;
 RA Kefelef B., Laforge K.S., Puigserver A., Scheele G.A.;
 RT "Primary structures of canine pancreatic lipase and phospholipase A2
 messenger RNAs.";
 RL Pancreas 1:430-437(1986).
 CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the
 2-acyl groups in 3-sn-phosphoglycerides.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O - 1-
 acylglycerophosphocholine + a fatty acid anion.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: D00035; BAA00023.1; -
 DR EMBL: M35301; AAA30883.1; -
 DR PIR: A24392; PSDG.
 DR PIR: S11316; S11316.
 DR HSP: P00592; 4P2P.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF00068; phoslip; 1.
 DR PRINTS: PR00389; PHPLIPASEA2.
 DR PRODOM: PD000303; PhospholipaseA2.
 DR SMART: SM00085; PA2c; 1.
 DR PROSITE: PS00118; PA2_HIS; 1.
 DR PROSITE: PS00119; PA2_ASP; 1.
 KW Hydrolase; Lipid degradation; Calcium; Pancreas; Signal.
 FT SIGNAL 1 15
 FT PROPEP 16 22
 FT CHAIN 23 146 PHOSPHOLIPASE A2.
 FT ACT_SITE 70 70 BY SIMILARITY.
 FT ACT_SITE 121 121 BY SIMILARITY.
 FT DISULFID 33 99 BY SIMILARITY.
 FT DISULFID 49 146 BY SIMILARITY.
 FT DISULFID 51 67 BY SIMILARITY.
 FT DISULFID 66 127 BY SIMILARITY.
 FT DISULFID 73 120 BY SIMILARITY.
 FT DISULFID 83 113 BY SIMILARITY.
 FT DISULFID 106 118 BY SIMILARITY.
 FT CA_BIND 50 50 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 52 52 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 54 54 VIA CARBONYL OXYGEN (BY SIMILARITY).
 FT CA_BIND 71 71 BY SIMILARITY.
 SQ SEQUENCE 146 AA; 16235 MW; F6258ED9527F3692 CRC64;

Query Match 12.3%; Score 70.5; DB 1; Length 146;
 Best Local Similarity 25.4%; Pred. No. 4.9;
 Matches 15; Conservative 11; Mismatches 18; Indels 15; Gaps 1;

OY 28 LDCEKTKCFKTEKCYVFEDYVYKECYVWVDEGLDQKCFVVDENFTEVYLTDEGKD 86
 DB 60 VDELDKCCQTHDHCY-----SEAKILDSCKFLDLNPRYKTYISYSCSGSE 103

RESULT 13
 ID CB20_HUMAN STANDARD; PRT; 156 AA.
 AC P52298; Q14924;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 20 kDa nuclear cap binding protein (NCBP 20 kDa subunit) (CBP20) (NCBP
 interacting protein 1) (NIP1).
 GN NCBP2 OR CBP20.
 OS Homo sapiens (Human).
 RN [3]

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBITaxID=9606;
 RN [1]
 RX MEDLINE=95379956; PubMed=7651522;
 RA Izaurralde E., Lewis J., Gamberl C., Jarmolowski A.,
 RA McGuigan C., Malta A.W.,
 RT "A cap binding protein complex mediating U snRNA export.";
 RL Nature 376:709-712(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=96017765; PubMed=7478990;
 RA Kataoka N., Ohno M., Koda I., Shimura Y.,
 RT "Identification of the factors that interact with NCBP, an 80 kDa
 RT nuclear cap binding protein.";
 RL Nucleic Acids Res. 23:3638-3641(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix;
 RA Strausberg R.,
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 FUNCTION: INVOLVED IN MEDIATING U SNRNA EXPORT FROM THE NUCLEUS.
 CC BINDS TO 5'CAPED MRNA.
 CC -1 SUBUNIT: THE CAP-BINDING PROTEIN (CBC) COMPLEX IS AN HETERODIMER
 CC OF CBP80 AND CBP20.
 CC -1 SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X84157; CAAS8962.1; -
 DR EMBL: D59253; BAA09599.1; -
 DR EMBL: BC001255; AAH01255.1; -
 DR HSSP: P11940; IGVJ.
 DR Genew; HGNC:7659; NCBP2.
 DR MIM: 605133; -
 DR InterPro: IPR005054; RNA_rec_mol.
 DR Pfam: PF00076; rrm.1.
 DR SMART: SMO0360; RRM.1.
 DR PROSITE: PS00103; RRM.1.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 DR Nuclear protein; RNA-binding.
 DR DOMAIN 40 118 RNA-BINDING (RRM).
 FT CONFLICT 97 97 A -> S (IN REF. 2).
 FT SEQUENCE 156 AA; 18001 MW; B6C94F3182A2CC3D CRC64;
 SQ
 Query Match 12.3%; Score 70.5; DB 1; Length 156;
 Best Local Similarity 21.8%; Pred. No. 5.2;
 Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;
 QY 1 EERE-----DCWTFYARKYTDYDKS-----FKKSSD-----LDDECKTKCFTEYCYIV 44
 DB 30 EDEQKLKRSCTLYVGNLSPTTTEQYLEFSSGDKTKITIMGLDKKTKA--CGFCFE 87
 QY 45 FEDTVNEKYVNVVGEELDEKFEVDENFTENTYLTDEGKDAGNAAGTDESDEVD 103
 DB 88 YSSADAEAMRYINGTRILDR--IIRTWDAGK---EGRYGRGRSGGVADERYQD 141
 RESULT 14
 Y875_METJA STANDARD; PRT; 748 AA.
 AC 058285;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0875.
 GN MJ0875.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanococcaceae; Methanococcococcus.
 NC NCBITaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Utechtack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67531; AAB98882.1; -
 DR TIGR: MJ0875; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 748 AA; 90731 MW; 5515ED3B541562EA CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 748;
 Best Local Similarity 24.4%; Pred. No. 27;
 Matches 29; Conservative 16; Mismatches 39; Indels 35; Gaps 6;
 QY 1 EEREDCWTFY-----ANRKYTDYDKSFKKSSDLDECKTKCFKTEYCYIV 47
 DB 144 KEFSDIWMYFDIKIYKKLSIKHRKEGDLDAIETL---KCYLAESYVKKFNDK 200
 QY 48 ----TYNKEKYVNVVGEELDEKFEVDENFTENTYLTDEGKDAGNAAGTDESDEVD 102
 DB 201 KKAEIFNKHRYNLM-----AKFESEKFEK-----AAEYKKRSGDTIKIDE 244
 RESULT 15
 S185_YEAST STANDARD; PRT; 1058 AA.
 AC P40856;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE S174-associating protein SAP185.
 GN SAP185 OR YJL098W OR J0840.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 NC NCBITaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96220458; PubMed=8649382;
 RA Luke M.M., della Seta F., di Como C.J., Sugimoto H., Kobayashi R.,
 RA Arndt K.T.,
 RT "The SAP, a new family of proteins, associate and function positively
 RT with the S174 phosphatase.";
 RL Mol. Cell. Biol. 16:2744-2755(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;

PX MEDLINE=95176706; PubMed=7871887;
RA Miosga T., Boles E., Schaaf-Gerstenschlaeger I., Schmitt S.,
Zimmermann F.K.;
RT "Sequence and function analysis of a 9.74 kb fragment of
RT Saccharomyces cerevisiae chromosome X including the BCK1 gene."
RL Yeast 10:1481-1488(1994).
RN [3]
RP SEQUENCE OF 1-853 FROM N.A.
RC STRAIN-S288c / FY1679;
RX MEDLINE=96090136; PubMed=7483851;
RA Rasmussen S.W.;
RT "A 37.5 kb region of yeast chromosome X includes the SMI1, MEF2, GSH1
RT and CSD3 genes, a TCP-1-related gene, an open reading frame similar
RT to the Dnl80 gene, and a tRNA(Arg).";
RL Yeast 11:873-883(1995).
CC -1- FUNCTION: ASSOCIATES WITH THE SIT4 PHOSPHATASE IN A CELL CYCLE
CC DEPENDENT MANNER. MAY BE DIRECTLY OR INDIRECTLY INVOLVED IN
CC SIT4-DEPENDENT FUNCTIONS IN BUDDING AND IN NORMAL G1 CYCLIN
CC EXPRESSION.
CC -1- PTM: HYPERPHOSPHORYLATED IN THE ABSENCE OF SIT-4.
CC -1- SIMILARITY: BELONGS TO THE SAPS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77923; CAA54892.1; -;
DR EMBL: X85021; CAA59396.1; -;
DR EMBL: Z49373; CAA89392.1; -;
DR SGD: S0003634; SAP185.
KW Phosphorylation; Cell cycle.
SQ SEQUENCE 1058 AA; 121402 MW; DD44D53DD87438 CRC64;

Query Match 12.2%; Score 70; DB 1; Length 1058;
Best Local Similarity 20.3%; Pred. No. 43;
Matches 25; Conservative 27; Mismatches 45; Indels 26; Gaps 5;

QY 2 ERECKWTFVYANRKYTDKSKSSDLDECKKTCFKTEY-----CYIYFE-- 46
Db 427 DRDPIYIGYLVKMFSEHMDFNKI--LITEKKIPLLOTSTGTIEPLGFERRKICELIAELL 484
QY 47 -----DTVNECYNYVVDGEELDQEKVVDENFTE-NYLDCGKDGAGNAAGTGESDEV 100
Db 485 HCSNMTLNLNPSAYDIYREDAERERLFPNSQNVDSNDRSELKENEDN---TGDADDEV 541
QY 101 DED 103
Db 542 EDD 544

Search completed: May 5, 2003, 17:49:27
Job time : 24 secs

Db 71 ENLAYCYSMNNNSKALRYFNLAALDNADIDCEKAI 107

RESULT 2

09X9S1 PRELIMINARY; PRT; 509 AA.

AC 09X9S1; 1

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Polysaccharide synthase.

GN TTS.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RA Llull D., Munoz R., Lopez R., Garcia E.;

RT "A single gene (tts) located outside the cap locus directs the formation of Streptococcus pneumoniae type 37 capsular polysaccharide: RT type 37 pneumococci are natural, genetically binary strains.";

RL J. Exp. Med. 19:241-251(1999).

DR EMBL: AJ131985; CAB51329.1;

DR InterPro: IPR001173; Glycos_transf_2.

DR Pfam: PF00535; Glycos_transf_2; 1.

SO SEQUENCE 509 AA; 58888 MW; C6C38D2ACFAB0D8A CRC64;

Query Match 13.7%; Score 78.5; DB 2; Length 509;

Best Local Similarity 31.1%; Pred. No. 9.8;

Matches 23; Conservative 10; Mismatches 30; Indels 11; Gaps 3;

OY 8 TFYANRKYTFDDKSKSSDDECKKTCFKTECYIVFEDTVN---KECYNVVVGEEELD 64

DB 74 TFYLNHYIHIGSSPKSSISEAKKVTL--YC-----TANFVECLVESMQDYAN 125

OY 65 OEKPVVDNFTENY 78

DB 126 FETVILDDSKSEVY 139

RESULT 3

09U2Q4 PRELIMINARY; PRT; 550 AA.

AC 09U2Q4; 1

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE Y26D4A.2 protein.

GN Y26D4A.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA White S.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL: AL110478; CAB54340.1; -

DR InterPro: IPR001304; Lectin_C.

DR SMART: SM00034; CLECT; 1.

SO SEQUENCE 550 AA; 62195 MW; 79AE220DB127466A CRC64;

Query Match 13.7%; Score 78.5; DB 5; Length 550;

Best Local Similarity 25.6%; Pred. No. 11;

Matches 23; Conservative 12; Mismatches 32; Indels 23; Gaps 3;

OY 6 CWTYANRKYTFD-----DKSFKS--SDDECKKTCFKTECYI 43

DB 170 CGNFMNRPYPDSCETECMDTWQVNAKPGPLMGSNINIDNLSSDECYKCYCMKDLNCFV 229

OY 44 VFEDTVNKECYNVVVDGEELDEKEFVDEN 73

DB 230 NYDKDKSKECQWMSIDNVHF-LEKVHPSN 258

RESULT 4

09XTS9 PRELIMINARY; PRT; 907 AA.

AC 09XTS9; 1

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Y69H2.2 protein.

GN Y69H2.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA McMurtry A.A.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL: Z98877; CAB11570.1; -

DR HSP: Q12784; 1HPE.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001458; EGF-11.

DR InterPro: IPR001873; Na+channel_ASC.

DR Pfam: PF00858; ASC; 1.

DR Pfam: PF00008; EGF; 7.

DR PRINTS: PR01078; AMINACHANNEL.

DR PRINTS: PR00010; EGFLOOD.

DR SMART: SM00181; EGF; 8.

DR PROSITE: PS00022; EGF_1; UNKNOWN_8.

DR PROSITE: PS01186; EGF_2; 4.

DR EGF-like domain; Glycoprotein.

SO SEQUENCE 907 AA; 101840 MW; 8B55FF139F8F81D7 CRC64;

Query Match 13.7%; Score 78.5; DB 5; Length 907;

Best Local Similarity 29.2%; Pred. No. 18;

Matches 33; Conservative 12; Mismatches 33; Indels 35; Gaps 8;

OY 10 YANRKYTFDDKSKF-----KSSD-----DECKKTC--FKTECYIVFEDT 48

DB 226 YAOEKSTD-DKTKLYTLGEGMOWSEDMVKDDVSMKDCSCGGDSYVC-----GK 279

OY 49 VNKECYNVVVDGEELDEKEFVDENFTENYLVNDCESKAGNAGTGESDEVD 101

DB 280 VNSRCI--VYENGEME-----TDPNACEDNSTLC-GAELGHGMCIWQSDVD 324

RESULT 5

09K8N5 PRELIMINARY; PRT; 682 AA.

AC 09K8N5; 1

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Two-component sensor histidine kinase involved in chemotaxis.

GN CHEA OR BH2970.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Bacillaceae; Bacillus.
 RX NCBI_TaxID=86665;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RA MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis";
 RT Nucleic Acids Res. 28:4317-4331 (2000).
 CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
 DR EMBL: AP001517; BAB06689.1; -;
 DR HSSP: 056310; 1830.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sens_pr_C.
 DR InterPro: IPR002545; Chem.
 DR InterPro: IPR004105; H-kinase_dlm.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR002570; Hpt.
 DR Pfam: PF01584; Chem; 1.
 DR Pfam: PF02895; H-kinase_dlm; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF01627; Hpt; 1.
 DR PRINTS: PR00344; BCTRLENSOR.
 DR ProDom: PD003142; Hpt; 1.
 DR SMART: SM00260; Chem; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00073; HPT; 1.
 DR Kase: Phosphorylation; Sensory transduction; Transferase;
 KW Complete proteome.
 SQ SEQUENCE 682 AA; 75272 MW; 46B61BE5BF7276 CRC64;

Query Match 13.6%; Score 78; DB 16; Length 682;
 Best Local Similarity 32.6%; Pred. No. 15;
 Matches 28; Conservative 13; Mismatches 31; Indels 14; Gaps 5;

QY 10 YANKRYTDPDKRFFKSSD-----LDECKTKCFTECYIYFE--DYVNECYIYVVG 60
 DB 156 YDEFTYVLDOSFEQGYQAYQLEVTLD--KTLKAARVMEVLEQVG-EVIKSPSA 212
 QY 61 EELDEKEFYVDENFENTENLTDEGKD 86
 DB 213 EELDEKEF--DERFLVTLTKVDGEE 236

RESULT 6
 ID 097239 PRELIMINARY; PRT; 3933 AA.
 AC 097239;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 467.9 kDa protein.
 GN PFC0245C, MAL3P2.18.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RA MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felwell T.,
 RA Genies S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
 RA Hurrell P., Jagsall K., Jassal B., Kyes S., McLean J., Moulé S.,
 RA Murrell K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skellern J., Squares R., Squares S., Sultson J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
 RT falciparum";
 RL Nature 400:532-538 (1999).

DR EMBL: AL034558; CAB39005.1; -;
 DR InterPro: IPR002048; EF-hand.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 3933 AA; 467876 MW; 51444604EE36933 CRC64;

Query Match 13.6%; Score 78; DB 5; Length 3933;
 Best Local Similarity 25.0%; Pred. No. 1e+02;
 Matches 23; Conservative 21; Mismatches 24; Indels 24; Gaps 5;

QY 36 FKTEYC-YIYF-----EFTVKECY---NVVDGEELD-QEKF-----VVD 71
 DB 1139 FFFNNCEYIIFLYNNHLLIEKEKLNKSCFYFLKNTLNCTCNKKNKFFWCFELPHIIR 1198
 QY 72 ENFENYLTDEGKADGNAAGTGESDEVED 103
 DB 1199 INEKSLKLNKTKIKRAGDDTDDDDDDDDDD 1230

RESULT 7
 ID 09VFH8 PRELIMINARY; PRT; 368 AA.
 AC 09VFH8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG14853 protein.
 GN CG14853.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moshrefi A.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasman D.A., Weisslock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;


```
QY      53 CYYNVVDGEELDQEK-VVDENFTE--NYLTDCGKDGAGNAAGTGDESDVD 101
      | : | | : ||| ::||::: | | : | | | : :||:
```

AF
SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Tortumi M., Wu H.C.,

RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamliya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narasaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Souhnick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At5g39780 (GI:15242471).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016876; BAB11383.1; -
 DR EMBL; AY070102; AAL49839.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 606 AA; 71043 MW; 5F02E244E59FC1SD CRC64;

Query Match 13.0%; Score 74.5; DB 10; Length 606;
 Best Local Similarity 27.7%; Pred. No. 29;
 Matches 31; Conservative 19; Mismatches 37; Indels 25; Gaps 7;

QY 1 EREEDCW--TFYKRRKTYD--KSFKSSDLDECKKCKTEKCYIVFDYVAKCY 55
 DB 103 ERTEDLGIVFHGEDKVKTEDYSSVSKK-----KKIRFLTE-----EDFLESDS-- 147
 QY 56 NVVDEEL---DQKFFVDENFTENYLTDCGKDGAGNAGTGDSEDEVD 103
 DB 148 DFVDSQFTTSNDEGFLSDSFATSLKKGONRSDN--SGSGSDSEEEEEE 198

RESULT 13
 ID 035821 PRELIMINARY; PRT; 1277 AA.
 AC 035821;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PAR interacting protein.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91084854; PubMed=2261643;
 RA Wharin J., Schibler U.;
 RT "Expression of the liver-enriched transcriptional activator protein
 RT DBP follows a stringent circadian rhythm.";
 RL Cell 63:1257-1266(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Comte P.A., Ossipow V., Schibler U.;
 RT "Isolation of PIP, a 160 kDa nuclear protein that interacts with the
 RT activation domain of PAR transcription factors.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U83590; AAB62878.1; -
 SQ SEQUENCE 1277 AA; 144675 MW; 43140A2973176792 CRC64;

Query Match 13.0%; Score 74.5; DB 11; Length 1277;
 Best Local Similarity 30.4%; Pred. No. 66;
 Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;

QY 1 EREEDCWTFYKRRKTYD--KSFKSSDLDECKKCKTEKCYIVFDYVAKCY 59
 DB 630 EDEED-----NVVYTDDEKOLKHEGADSDS-----EDSKNSCEDVSED 670
 QY 60 GEELDOE--KFFVDENFTENYLTDCGKDGAGNAGTGDSEDE 99
 DB 671 GESEEDRDKVDVDPGFRQOLM---EVLQAGNALGGESEEE 709

RESULT 14
 ID 09FK41 PRELIMINARY; PRT; 596 AA.
 AC 09FK41;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similarity to unknown protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Sequence features of the regions of 1,367,185 bp covered by 19
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:203-216(1998).
 DR EMBL; AB012248; BAB09450.1; -
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF01344; Kelch; 6.
 SQ SEQUENCE 596 AA; 68483 MW; 8865D3CAD873CBD0 CRC64;

Query Match 12.9%; Score 74; DB 10; Length 596;
 Best Local Similarity 24.2%; Pred. No. 32;
 Matches 29; Conservative 23; Mismatches 40; Indels 28; Gaps 5;

QY 12 NRKYTDFDKSKF-----KSSDLDECKKCKTEKCYIVFDYV-----NKEC 53
 DB 339 NRRMSDKSKALQAKRTLPQVSSEVVKPCGRINSCMVGVGDTLYIGCWMEIDKEV 398
 QY 54 YYN---VVDGEELDQKFFVDENFT--NYLTDCG-----KAGNAGTGDSEDEVD 103
 DB 399 TLDDIVSLNLSKIDEMWCIPTTETEWVEVSDEGDEDDSDSEDESEDEDD 458

RESULT 15
 ID 050667 PRELIMINARY; PRT; 1278 AA.
 AC 050667;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein BBH09.
 GN BBH09.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Gasjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Cocayne J.D., Weidman J.,
 RA Ungerback T., Wathley L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586(1997).
 DR EMBL; AE000784; AAC66000.1; -
 DR TIGR; BBH09; -
 DR InterPro; IPR002296; NI2N6_mtfase.
 DR InterPro; IPR002052; N6_mtfase.
 DR PRINTS; PR00507; NI2N6MTFRASE.
 DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
 KW Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 1278 AA; 150871 MW; 8A0B22DF16565C0 CRC64;

Query Match 12.9%; Score 74; DB 16; Length 1278;

Tue May 6 16:59:49 2003

us-09-936-737a-2.rsp

Page 7

Best Local Similarity 25.0%; Pred. No. 74;
Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

```
OY      10 YANKRYIDFDSKSPFKSSDLBCKCTKTERTCYIVFEEDJVNKECYNVW-----VDGEILD   64  
        |::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     181 YGCRKYIIFD--FSKIREEK----YYEOWFIILFYLRIRERYLTNISIISVEKEQIS    233
```

```
QY      65 QEFVVDENFTE 76
        :||:::|
Db      234 KEKEIQKTLRE 245
```

```
Search completed: May 5, 2003, 17:52:37
Job time : 71 secs
```

